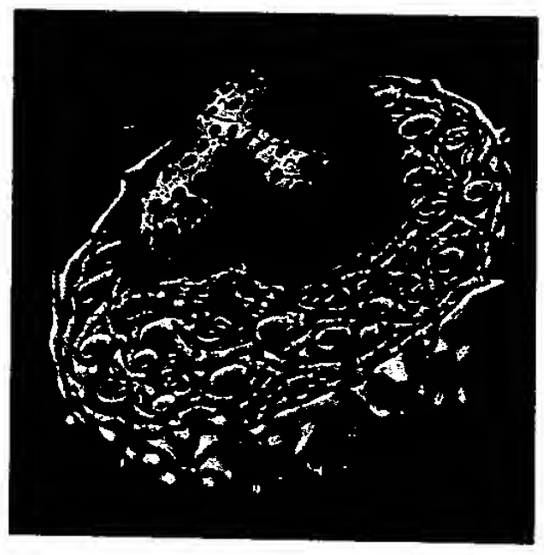
Molecular Cell Biology

SECOND EDITION



JAMES DARNELL

Vincent Astor Professor
Rockefeller University

▲

HARVEY LODISH

Member of the Whitehead Institute for Biomedical Research
Professor of Biology, Massachusetts Institute of Technology

▲

DAVID BALTIMORE

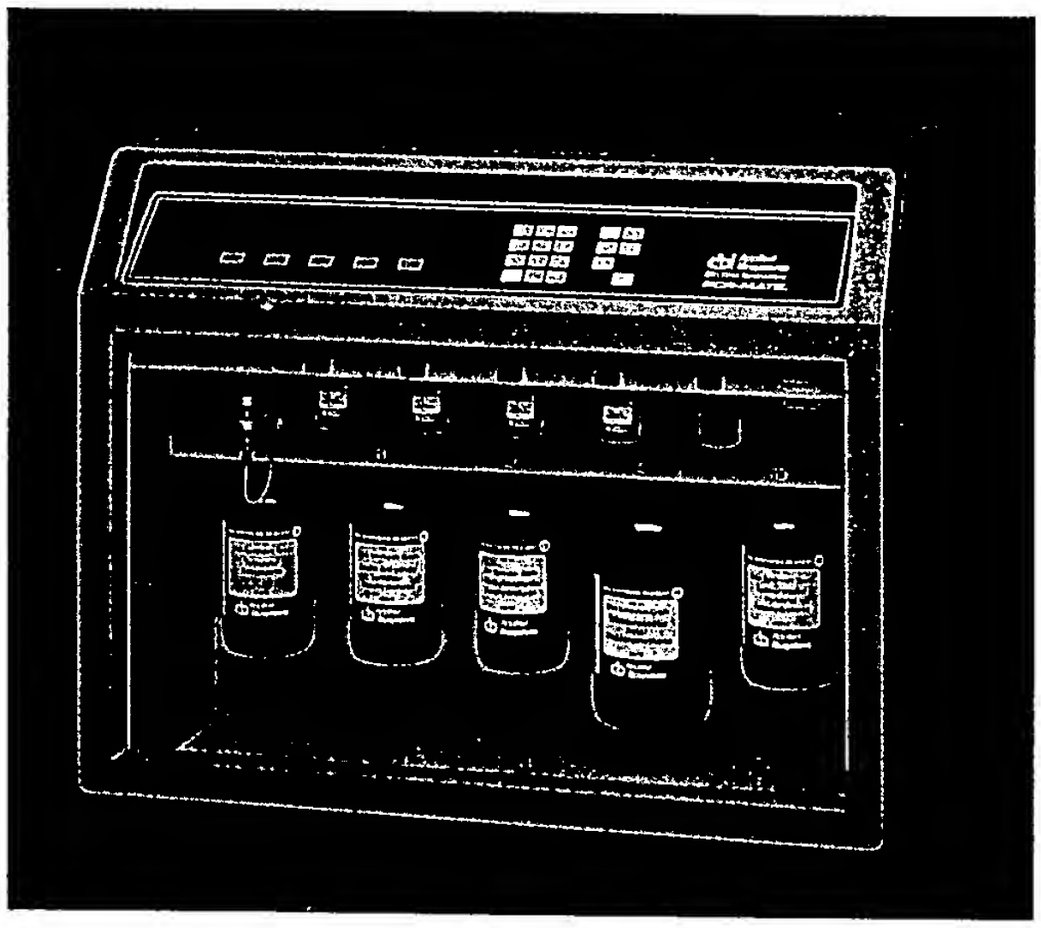
President
Rockefeller University

SCIENTIFIC AMERICAN BOOKS

Distributed by W. H. Freeman and Company, New York

CHAPTER 6

Manipulating Macromolecules



DNA synthesizer

The greatest advances in molecular cell biology in the recent past have been based on the analysis and manipulation of macromolecules, particularly DNA. For years it was clear that many deep biological secrets were locked up in the sequence of bases in DNA, but obtaining the sequences of long regions of DNA—not to mention altering these sequences at will—seemed a distant dream. An avalanche of technical advances in the 1970s drastically changed this perspective. First, enzymes were discovered that cut the DNA from any organism at specific short nucleotide sequences, generating a reproducible set of pieces. The availability of these enzymes, called restriction endonucleases, *greatly facilitated two important developments: DNA cloning and DNA sequencing.*

Two DNA molecules can be joined enzymatically and thus restriction fragments of any DNA can be inserted into a variety of vectors, often plasmid DNA, to produce recombinant DNA. *The recombinant molecules can be introduced into an appropriate cell population (most often bacteria) and cells containing recombinant DNA*

molecules can be selected. This procedure is referred to as cloning *the chosen DNA sequence. Once a clone of cells bearing the recombinant DNA is selected, unlimited quantities of the chosen DNA can be prepared. In addition, DNA oligonucleotides (up to 100 bases long) can now be chemically synthesized entirely automatically. Not only natural oligonucleotides, but any desired mutant sequence can be produced and can be inserted in recombinant DNA.*

Rapid DNA sequencing *came into being in the late 1970s. By the use of restriction endonucleases, long DNA molecules from a single organism could be broken into a reproducible array of fragments, whose order in the original molecule could then be determined. It also became possible to determine the sequence of bases in fragments containing as many as 500 nucleotides. There was no longer any obstacle to obtaining the sequence of a DNA molecule containing 10,000 or more nucleotides. Suddenly, any DNA was accessible to isolation and to sequencing. Now, computer-automated procedures for sequencing and for storing, comparing, and analyzing data will make it possible to obtain the entire sequence of the human genome in the near future.*

Any desired DNA, whether natural, modified, or completely synthetic, can be reinserted into cells and tested for biological activity. Selected DNA fragments that encode proteins of particular interest have been transferred to bacteria and to other cells, where the transferred DNA directs the production of natural or mutant proteins. In addition, the direct chemical synthesis of peptides as much as 75 amino acids long is also now routine.

Almost overnight, this group of techniques, often collectively called molecular genetics, *became the dominant approach to the study of many basic biological questions, including how gene expression is regulated in eukaryotic cells and how protein or domains of proteins function. The power and success of the new technology have raised high hopes that the practical use of our ever-increasing biological knowledge will bring many benefits to mankind.*

This chapter outlines the techniques just summarized and describes some older procedures still widely used in molecular experiments today. ▲

Radioisotopes: The Indispensable Modern Means of Following Biological Activity

A major goal of classic biochemistry from the 1930s through the 1950s was to chart the metabolic pathways in cells. This work contributed enormously to our present detailed picture of cellular biochemistry. Almost all this fundamental biochemical knowledge was garnered with chemical or enzymatic assays that relied on simple tests using color indicators or measurements of light absorption at characteristic wavelengths.

Since World War II, when radioactive materials first became widely available as byproducts of work in nuclear physics, chemists and biologists have fashioned an almost limitless variety of radioactive "tracer" molecules. Today, the radioactively labeled precursors of macromolecules greatly simplify many standard biochemical assays and significantly enhance our ability to follow biochemical events in whole cells. Almost all experimental biology depends on the use of radioactive compounds.

In a labeled molecule, at least one atom is present as a *radioisotope* (in radioactive form; see Table 6-1). This radioisotope does not change the chemical properties of the molecule. For example, an enzyme—whether it is in a cell extract or a living cell—does not distinguish between a labeled and an unlabeled molecule when performing a metabolic function, such as synthesizing protein, DNA, or RNA. Because radioactive atoms can be detected when they emit a particle, they can be used to trace the activities of the labeled molecules.

Not all labeled materials can be used interchangeably in whole cells and cell-free systems. For one reason, many compounds that participate in intermediary metabolism cannot be used to study the metabolism of whole cells because they do not enter cells. For example, labeled ATP may contribute phosphorus 32 (^{32}P) in RNA synthesis in a cell-free system, but ATP does not enter cells. On the other hand, labeled orthophosphate ($^{32}PO_4^{3-}$) added to the medium does enter cells and is incorporated into nucleotides and then into cellular RNA; but $^{32}PO_4^{3-}$ is not efficiently incorporated into nucleotides in cell extracts and therefore is not used to label RNA in extracts.

Amino acids and nucleotides labeled with either carbon 14 (^{14}C) or tritium (^{3}H) are commercially available, as are hundreds of labeled metabolic intermediates. Methionine labeled with sulfur 35 (^{35}S) is widely used as a protein label because it is available in high specific activities. The *specific activity* is the amount of radioactivity per unit of material; for example, commercial [^{35}S]-methionine can have over 10^{15} disintegrations per minute per millimole of methionine. The magnitude of the specific activity depends on the ratio of unstable (potentially radioactive) atoms to stable (nonradioactive) atoms and on the probability of decay of the unstable atoms, indicated by their half-life.

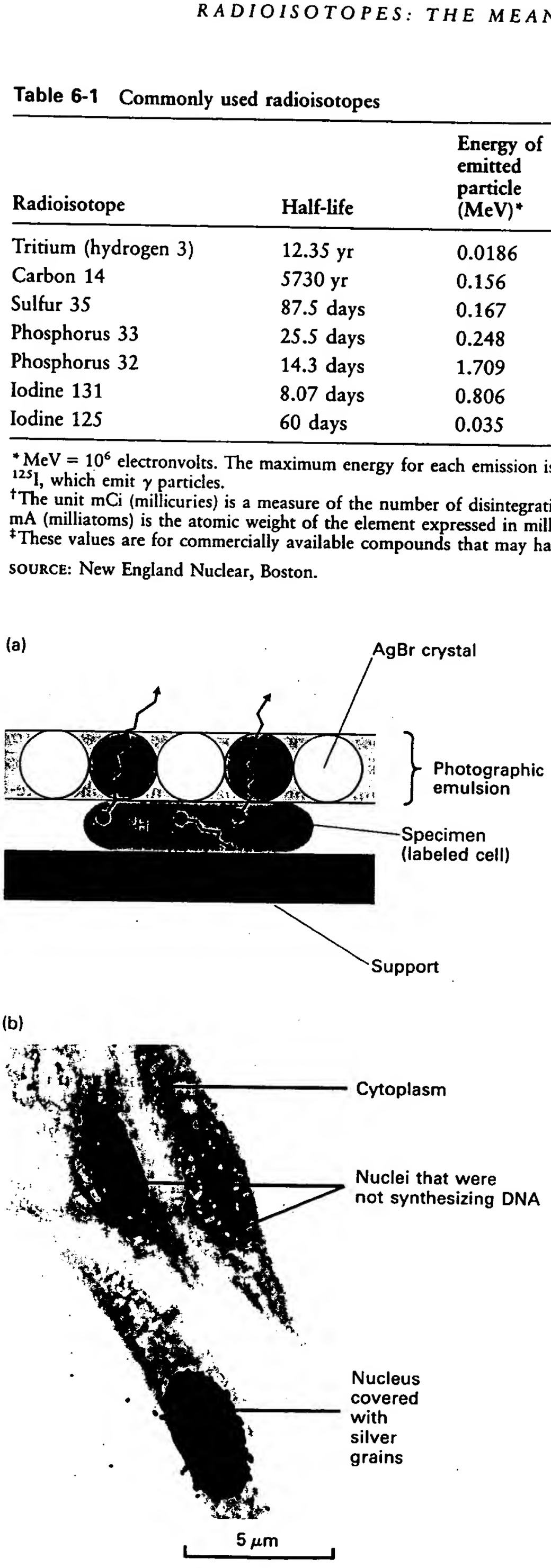
Because phosphate in which every phosphorus atom is ^{32}P is readily obtainable, various cell-free methods of labeling nucleic acids employ nucleotides labeled with ^{32}P. Likewise, a radioactive isotope of iodine (^{125}I) is available in almost pure form; this tracer atom can be enzymatically or chemically attached to a protein or nucleic acid without drastically affecting the macromolecule.

Table 6-1 Commonly used radioisotopes

Radioisotope	Half-life	Energy of emitted particle (MeV)*	Mean path length in water (μm)	Specific activity (mCi/mA)†	Common specific activities for compounds (mCi/mmol)‡
Tritium (hydrogen 3)	12.35 yr	0.0186	0.47	2.92×10^4	$10^2 - 10^5$
Carbon 14	5730 yr	0.156	42	62.4	$1 - 10^2$
Sulfur 35	87.5 days	0.167	40	1.50×10^6	$1 - 10^6$
Phosphorus 33	25.5 days	0.248	—	5.32×10^6	$10 - 10^4$
Phosphorus 32	14.3 days	1.709	2710	9.2×10^6	$10 - 10^5$
Iodine 131	8.07 days	0.806	—	1.6×10^7	$10^2 - 10^4$
Iodine 125	60 days	0.035	—	2.2×10^6	$10^2 - 10^4$

*MeV = 10^6 electronvolts. The maximum energy for each emission is given. The particle emitted is a β particle, except in the case of ^{131}I and ^{125}I, which emit γ particles.
†The unit mCi (millicuries) is a measure of the number of disintegrations per time unit: 1 mCi = 2.2×10^9 disintegrations per minute. The unit mA (milliatoms) is the atomic weight of the element expressed in milligrams.
‡These values are for commercially available compounds that may have many carbon or hydrogen atoms.

SOURCE: New England Nuclear, Boston.



Radioisotopes Are Detected by Autoradiography or by Quantitative Assays

Two detection schemes for assaying incorporated radioactivity are in general use:

1. In *autoradiography,* a cell or cell constituent is labeled and then overlaid with a photographic emulsion sensitive to radiation. Development of the emulsion reveals the distribution of labeled material. In whole cells, autoradiographic studies determine the original sites of the synthesis of macromolecules and their subsequent movements within cells. For example, incorporation of [^{3}H]thymidine identifies the nucleus as the major site of DNA synthesis and cell fractionation and histologic staining show most DNA is also in the nucleus (Figure 6-1). In contrast, the incorporation of

◀ **Figure 6-1** The technique of autoradiography. (a) A radiation-sensitive photographic emulsion containing silver salts (AgBr) is placed over labeled cells attached to a glass slide (for the light microscope) or to a carbon-coated grid (for the electron microscope). The cell regions containing the labeled molecules emit radioactive particles, along the tracks of which silver is deposited. When the photographic emulsion is developed, the silver deposited appears as dark grains under the light microscope and as curly filaments in the electron microscope. (b) These fibroblasts from Chinese hamsters were labeled with [^{3}H]thymidine for 1 h. Two of the cells were not synthesizing DNA during this time (the larger dark areas in their nuclei are nucleoli), but one cell was. Small black grains almost entirely cover that cell nucleus, indicating the new DNA is there. *Part (a) redrawn from E. D. P. DeRobertis and E. M. F. DeRobertis, 1979,* Cell and Molecular Biology, *Saunders, p. 62; part (b) courtesy of D. M. Prescott.*

labeled uridine into RNA has shown that most RNA is first made in the nucleus but that most RNA in cell fractions is located in the cytoplasm. Incorporation of labeled amino acids has revealed that most protein is made in the cytoplasm. The transport pathway of proteins from synthesis to secretion was first documented by electron microscopic autoradiography, which allows each silver filament that results from a radioactive disintegration to be observed.

2. In *quantitative assays,* cells are labeled either in vivo or in vitro and their constituents are isolated and purified in various ways. The amount or type of radioactivity in these constituents is then measured—by a Geiger counter, which detects ions produced in a gas by the radioactive emissions, or by a scintillation counter, which counts the flashes of light generated by mixing the radioactive sample with a substance that fluoresces after absorbing the energy of a particle resulting from the decay of the nucleus of the radioactive atom.

A combination of labeling and biochemical techniques is often employed. A cell constituent may be purified before it is labeled and, after labeling, be subjected to experimental procedures. Autoradiography of the labeled products of such experiments—most often after they have been separated by gel electrophoresis (discussed later in this chapter) or by chromatography—is perhaps the most common experiment in all of modern biology.

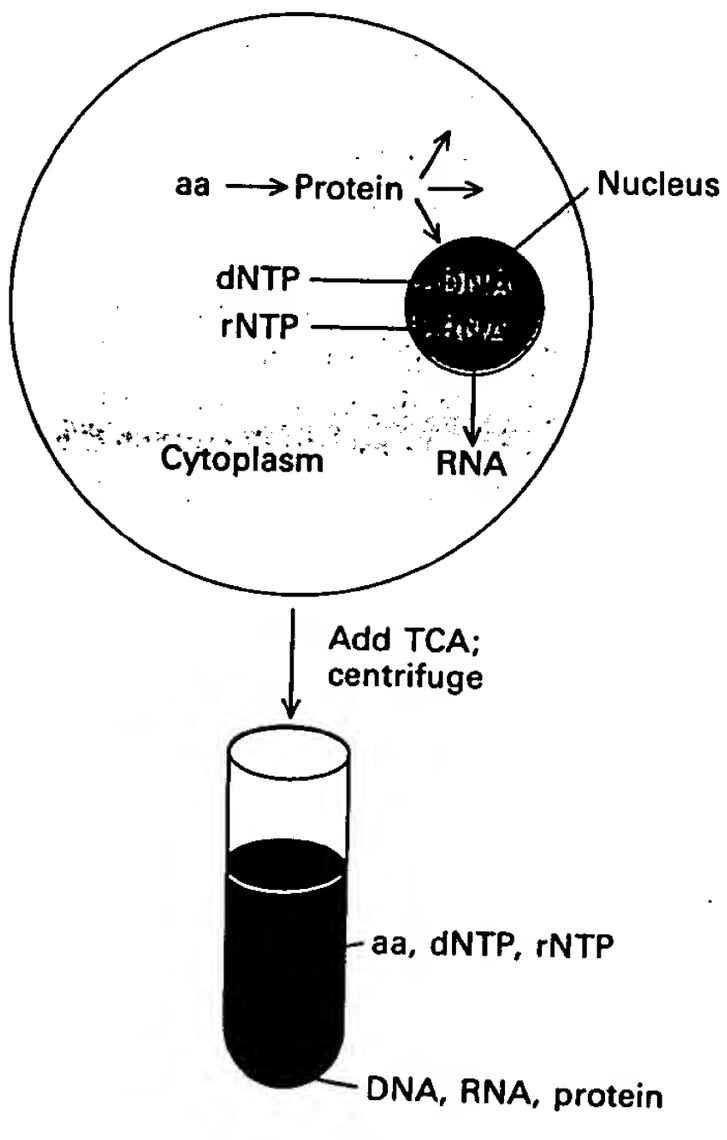
The purpose of the experiment governs the choice of a radioisotope as well as the detection method. A labeled compound must have a high enough specific activity that the radioactivity in the cell fraction is significant enough to be studied when the compound is incorporated into cells. For example, ^{3}H-labeled nucleic acid precursors are available in much higher specific activities than ^{14}C-labeled samples are; the former allows RNA or DNA to be adequately labeled after a shorter time of incorporation or in a smaller cell sample.

In autoradiographic studies, the energy in the particles released by radioactive disintegrations affects the experimenter's ability to localize the site at which the radioactivity is incorporated. For example, the β particles emitted by ^{32}P are so energetic (see Table 6-1) that the streaks they make on photographic film can be as long as 1 mm, much longer than the diameters of individual cells. ^{3}H is highly preferred for locating radioactive substances or structures in cells: the track created on photographic film by the β particle released by ^{3}H decay is only about 0.47 μm long; thus ^{3}H-labeled structures can be located within cells to an accuracy of about 0.5–1.0 μm, or about one-fifth the diameter of the nucleus of a mammalian cell.

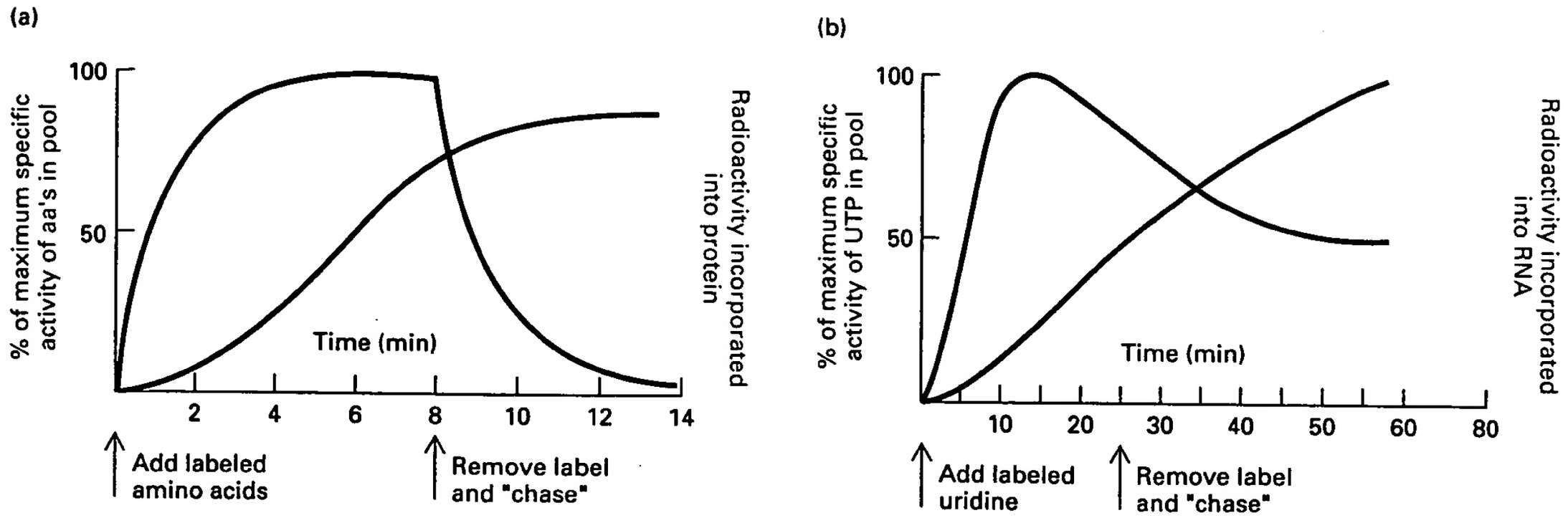
Pulse-Chase Experiments Must Be Designed with Knowledge of the Cell's Pool of Amino Acids and Nucleotides

In many experiments using radioactive metabolic material, a labeled compound is added to cells and the path of the labeled compound can then be traced as it moves through various compartments or molecules within cells. One type of experiment, the *pulse-chase experiment,* utilizes the brief addition (a *pulse*) of a labeled compound, followed by its removal and replacement (the *chase*) by an excess of unlabeled compound; the cells or cell constituents are examined at various times thereafter to monitor the radioactivity incorporated during the pulse.

Before an amino acid, a nucleoside, or a phosphate ion (for example) is incorporated into a protein or a nucleic acid, it enters the cell's *pool* of molecular building blocks—a collection of small molecules free to diffuse throughout the cytoplasm and nucleus of the cell but not necessarily free to diffuse into or out of membrane-bound organelles,



◀ **Figure 6-2** A cell's pool of small soluble molecules—amino acids (aa) and nucleotides (dNTP and rNTP)—may be separated from the macromolecules (DNA, RNA, and proteins) by adding cold acid, usually trichloroacetic acid (TCA), which destroys the cell structure and precipitates all macromolecules. Centrifugation then deposits the macromolecules in a pellet, leaving the amino acids and nucleotides in the supernatant. The rate at which cells take up labeled molecules and incorporate them into macromolecules can be determined by taking such samples at frequent intervals after the addition of labeled amino acids or nucleotide precursors to the cell-culture medium.



▲ **Figure 6-3** (a) If growing cells are exposed to a medium containing labeled amino acids, it takes about 5 min for the amino acids in the cell pool to reach the maximum specific activity. The accumulation of radioactivity in proteins starts more slowly, because the label must be incorporated into the amino acid pool first. However, if medium containing unlabeled amino acids is used instead, the incorporation of radioactivity into proteins stops within a few minutes due to the rapid equilibration between amino acids inside the cells and in the medium. Thus a marked pulse-chase effect is seen. (b) A pulse of labeled uridine is incorporated into UTP in the cell pool in about 10 min, and the pool can be diluted somewhat by excess unlabeled nucleosides outside the cells (drop in specific activity at 25 min). However, it takes much longer for a chase with unlabeled uridine to level off the amount of radioactivity incorporated into RNA (reduce the ongoing incorporation of radioactivity to 0). This is because the labeled uridine in the pool has been phosphorylated and is unable to escape to the medium, and there are almost 20 percent more uridine nucleotides in the pool than in cellular RNA. All of the labeled uridine is eventually incorporated into the RNA, but only after several hours. Thus no marked pulse-chase effect is seen.

such as the mitochondria or chloroplasts. Depending on the growth conditions of the cell the quantities of components of the pool can vary. Likewise, the rates at which different molecules are absorbed, utilized, and secreted by the cell can also vary (Figure 6-2).

Because of the rapid exchange of amino acids between the pool and the medium, a clear pulse-chase effect can be achieved with them. The acid-soluble pool can be made to contain radioactive amino acids in a few seconds, and they can be removed just as quickly (Figure 6-3a).

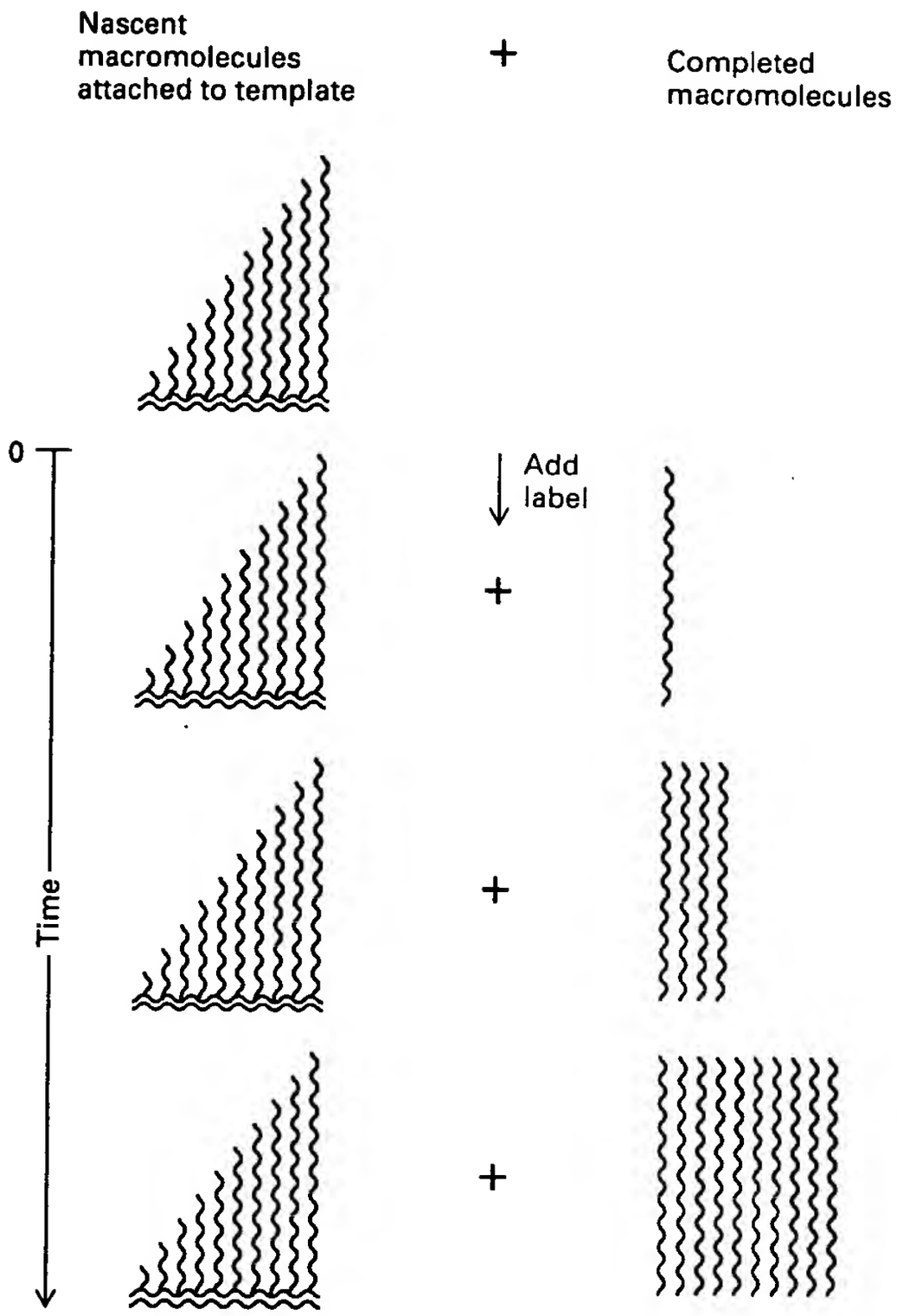
Ribonucleosides and deoxyribonucleosides, however, become phosphorylated soon after they enter the cell pool, and phosphorylated compounds do not generally leave the cell. Thus labeled nucleosides can enter the cell, but no equilibrium is established between the nucleic acid precursors in the medium and their phosphorylated counterparts in the cell. Nevertheless, a practically useful pulse-chase effect can be obtained in experiments with radioactive *deoxyribonucleosides,* because the deoxyribonucleotide content of the cell pool is sufficient for only a few minutes of DNA synthesis. Labeled thymidine, for example, can be satisfactorily chased even though it is phosphorylated, because an amount of thymidylate (TTP) equal to that in the pool is taken up every few minutes by replicating DNA.

Labeled *ribonucleosides* behave differently, because it takes several hours for enough RNA synthesis to occur to consume the content of the ribonucleotide pool in animal cells. In most cultured animal cells, the pool does absorb a pulse of labeled ribonucleosides quickly, say within 10 min. However, a marked chase response (one that occurs within a few minutes) is not possible. Although the addition of unlabeled ribonucleosides to the exterior medium may further expand the ribonucleotide content of the cell pool, dilute the label within it, and decrease the rate of RNA labeling, the amount of incorporated label does not clearly level off until several hours after the chase begins (Figure 6-3b).

In planning and interpreting experiments that use labeled precursors of proteins, DNA, or RNA to study macromolecular synthesis, these characteristics of small molecules in the soluble pool must always be borne in mind.

Labeled Precursors Can Trace the Assembly of Macromolecules and Their Distribution in a Cell

When a radioactive building block first enters a cell, it can only label the macromolecules that are in the process of being constructed. For example, if a radioactive amino acid is added to a culture, the nascent (unfinished and still growing) protein chains are the first proteins to be la-



▲ **Figure 6-4** Labeled radioactive precursors (red) first appear in nascent macromolecules. As time passes, molecules that contain more radioactive label are completed. At the end of an interval equivalent to the synthesis time of a macromolecule, the total amounts of radioactivity in all finished and all unfinished molecules are equal.

beled. As time passes, an increasing number of completed chains contain the radioactive label. The time required to form a specific macromolecule can be estimated by sampling a labeled cell culture at very short intervals to compare the amount of radioactivity in all nascent macromolecules still attached to the templates with the amount in all free (complete) macromolecules. The first finished chains obtained after the label is added contain only a small amount of the label, because they were almost completed before it was introduced. Each nascent chain initially also contains a small amount of label; as time passes, however, more label accumulates in newly finished chains and the nascent chains become completely labeled. At this point, there is an equal amount of label in the finished and nascent chains (Figure 6-4). The time elapsed since the label was added is equal to the time required for the synthesis of one chain.

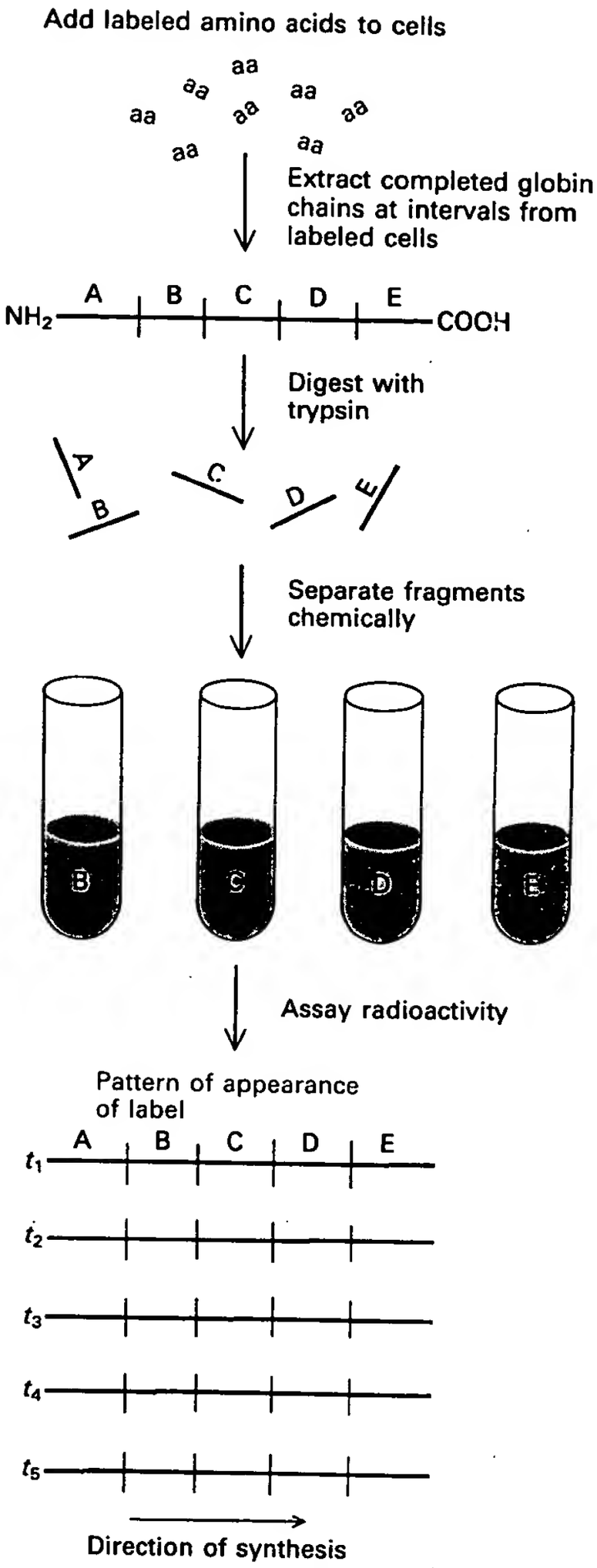
The Dintzis Experiment Proved That Proteins Are Synthesized from the Amino End to the Carboxyl End Other very important facts—the cellular locus of synthesis of a macromolecule and the direction of its growth—can be determined by labeling growing chains. Indeed, the analysis of newly finished chains was used by Howard Dintzis in a classic experiment demonstrating the step-by-step formation of protein chains from the amino terminus to the carboxyl terminus. Over 90 percent of the protein synthesized by *reticulocytes* (the next-to-final stage in the differentiation of red blood cells in the bone marrow of mammals) consists of the α- and β-globin chains that form the protein part of hemoglobin. (Hemoglobin is composed of four globin chains: two α and two β.) Dintzis exposed reticulocytes to radioactive amino acids and then, at short intervals, collected the finished chains. He separated the α and β chains and digested each with trypsin, an enzyme that attacks on the carboxyl side of arginine and lysine residues to produce a specific set of fragments for each chain which can be separated. Dintzis knew the sequence of amino acids in both globin chains as well as the position of each fragment within the globin chains.

Dintzis reasoned that the first completed chains to contain the radioactive label would be those that were almost complete when the label was added. Thus the *first* portion of the finished chains to contain label would be near the end at which chain synthesis finished and, by extension, the *last* portion of the finished globin chains to become labeled would lie at the end where chain synthesis started. The results showed that the radioactive label always appeared in the tryptic fragments in a certain order—in the carboxyl-terminal fragment first and the amino-terminal fragment last, with intermediate fragments becoming consecutively labeled in the order in which they lay between the two termini (Figure 6-5). From this, Dintzis deduced that synthesis begins at the amino terminus of each chain and moves in a step-by-step progression to the carboxyl end of the chain.

Whereas Dintzis studied the labeling of newly finished molecules, other workers have studied nascent molecules. (Experiments on the labeling of nascent RNA and DNA are described in Chapters 8 and 12.) The logic of these studies parallels that of the Dintzis experiment: the shortest labeled molecules in a nascent set will be those whose sequence is near the start site; increasingly longer members will contain additional sequences progressively more remote from the start site.

Determining the Sizes of Nucleic Acids and Proteins

Whereas the *sequence* of the monomers in a protein or nucleic acid ultimately determines the functional capacity of the polymer, the most useful physical characteristic in the



▲ **Figure 6-5** A diagram of Howard Dintzis' classic experiment showing the growth of polypeptides from amino to carboxyl end. Soon after radioactively labeled amino acids were added to a suspension of reticulocytes, finished labeled globin chains were released from ribosomes. After labeling, samples of released chains were taken at frequent intervals and cleaved into fragments by digesting the protein with the protease trypsin; radioactivity in the fragments was then assayed. In the first sample of finished chains (t_1), all radioactivity (red) was located in the E fragments at the carboxyl end. In samples t_2–t_5, fragments were labeled in the order D, C, B, and finally A. The direction of synthesis was therefore A → E. Thus Dintzis concluded that protein synthesis begins at the amino terminus and progresses to the carboxyl terminus. [See H. Dintzis, 1961, *Proc. Nat'l Acad. Sci. USA* 47:247.]

analysis of a polymer is its unique length. It is relatively easy to separate molecules by length; it is more difficult (for proteins) or virtually impossible (for nucleic acids) to separate molecules by using chemical differences based on sequence differences. Therefore, the length (size) of a protein, RNA, or DNA molecule is one of the most frequent measurements in molecular cell biology. In the following sections, we briefly outline the principles of molecular separation according to size and illustrate their use. The newer techniques are so simple and effective that they may not be appreciated as "physics in action." Students are encouraged to learn the principles of physical chemistry that underly these crucial techniques (see the references for this section, particularly those of Cantor and Schimmel).

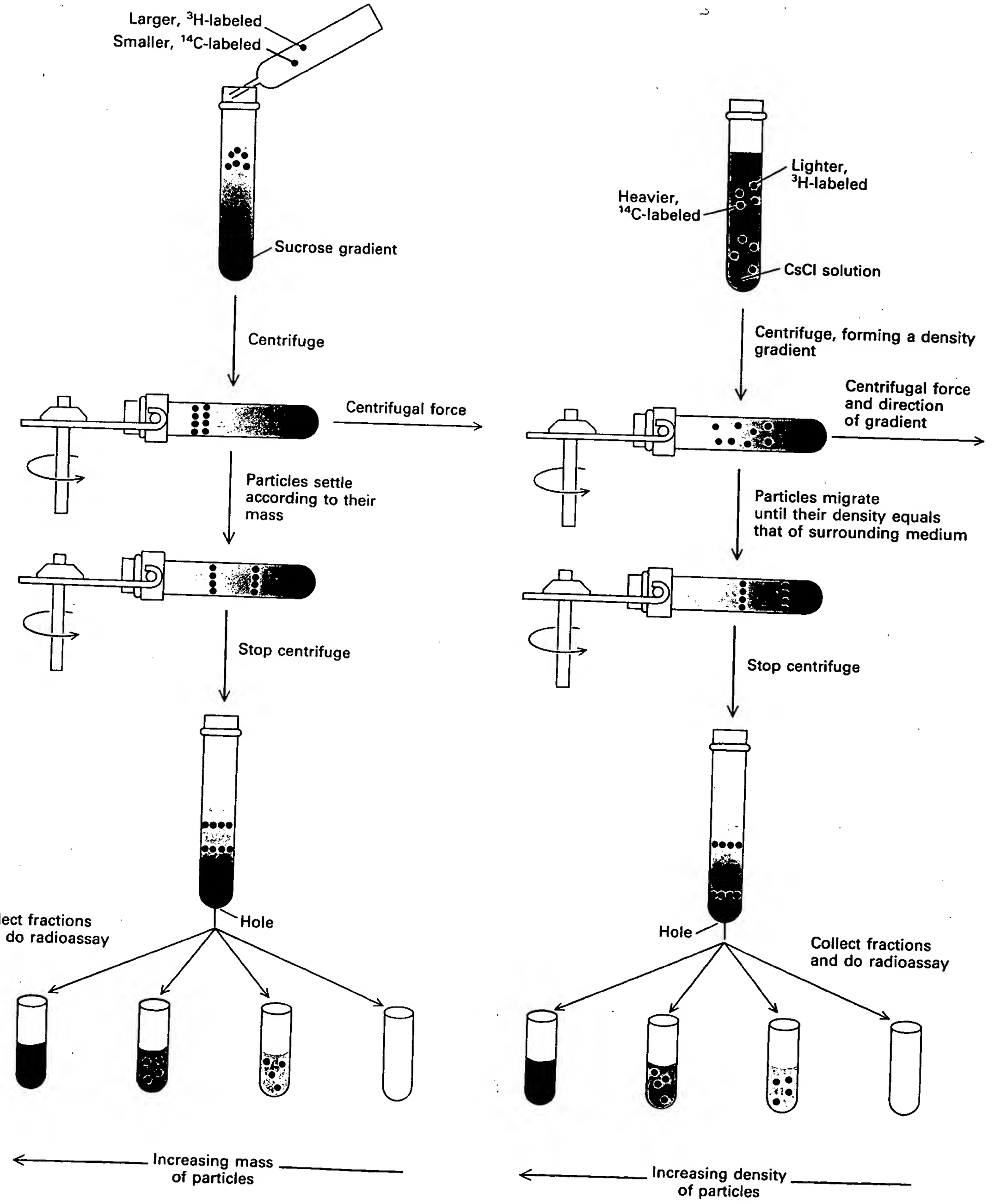
Centrifugation Is Used to Separate Particles and Molecules That Differ in Mass or Density

Two basic uses of centrifugation recur in the experiments described in this book: (1) separation of particles according to their mass, and (2) separation according to their density. We shall discuss each use in turn.

Rate-Zonal Centrifugation When particles or molecules are layered on top of a liquid column in a tube and subjected to centrifugation, they migrate down the tube at a rate controlled by the centrifugal force, the mass of the particles, the difference between the densities of the particles and the suspending medium, and the friction between the particles and the suspending medium (which depends on the shape of the particles). For example, RNA molecules of similar average shape and density separate in a centrifugal field almost solely according to mass, or chain length. After centrifugation is complete, different-sized molecules are found in different zones of the centrifuge tube; this separation technique is commonly called *rate-zonal centrifugation* (Figure 6-6a). Samples are centrifuged just long enough to separate the molecules of interest. If they are centrifuged for too short a time, the molecules will not separate sufficiently; if they are centrifuged much longer than necessary, all of the molecules will end up in a pellet at the bottom of the tube.

To prevent stirring the contents of the centrifuge tube during acceleration and deceleration, the liquid column through which particles are sedimented is often stabilized by a sucrose solution that is more concentrated (and thus denser) at the bottom than the top of the tube. For this reason, the technique is sometimes called "sucrose density-gradient centrifugation," but this terminology is erroneous because the property mainly responsible for the separation of particles by rate-zonal centrifugation is not density, but *mass*.

Although the sedimentation rate is strongly influenced by particle mass, this technique is seldom effective in de-


(a) RATE-ZONAL CENTRIFUGATION
Sample is layered on top of gradient
Larger, ³H-labeled
Smaller, ¹⁴C-labeled
Sucrose gradient
Centrifuge
Centrifugal force
Particles settle according to their mass
Stop centrifuge
Collect fractions and do radioassay
Hole
Increasing mass of particles
(b) EQUILIBRIUM CENTRIFUGATION
Sample is mixed with CsCl solution
Heavier, ¹⁴C-labeled
Lighter, ³H-labeled
CsCl solution
Centrifuge, forming a density gradient
Centrifugal force and direction of gradient
Particles migrate until their density equals that of surrounding medium
Stop centrifuge
Hole
Collect fractions and do radioassay
Increasing density of particles

◀ **Figure 6-6** Two centrifugation techniques are widely used to separate different subcellular particles, different types of molecules, or the same type of molecules of different mass. When they have separated sufficiently, the centrifuge is stopped and samples are collected from a hole punctured in the bottom of the tube. The different samples may be identified by various assays. (a) *Rate-zonal centrifugation* separates particles or molecules that differ in mass but may be similar in shape and density (for example, RNA molecules). Here, two particles of different mass have been labeled before centrifugation. (b) *Equilibrium centrifugation* allows the separation of particles that differ in density (for example, because they contain different ratios of protein, DNA, or RNA). The particles may or may not differ in mass and shape. Particles move up or down through a density gradient established when the centrifugal force acts on a dissolved salt such as CsCl. At equilibrium, the particles in the tube collect at levels at which the density of the solution equals their own density.

termining *exact* molecular weights because variations in shape also affect sedimentation rate. The exact effects are hard to assess, especially for proteins and single-stranded nucleic acid molecules that can assume many complex shapes. Nevertheless, rate-zonal centrifugation has proved to be the most practical method for separating many different types of polymers and particles.

While an analytical ultracentrifuge (one equipped with optical instruments that can record light absorption in the ultraviolet range, for example, where nucleic acids absorb strongly) is in motion, the sedimentation rate can be measured by photographing the moving boundaries of sedimenting layers of molecules. Modern ultracentrifuges reach speeds of 60,000 revolutions per minute (r/min) or greater and generate forces sufficient to sediment particles with masses greater than 10,000 daltons. For a particle located 6 cm from the rotational axis of a centrifuge, 60,000 r/min corresponds to a centrifugal force of 250,000 times gravity (250,000*g*). However, even at such tremendous forces, quite small particles with masses of 5000 daltons or less diffuse too freely to settle uniformly through a centrifugal field.

Equilibrium Density-gradient Centrifugation A density gradient can be established throughout the suspending medium before centrifugation. Alternatively, the force of centrifugation itself can be used to establish a density gradient; this separation technique is called *equilibrium centrifugation* (Figure 6-6b). In both cases, the density of the medium should range from less dense than the particles to be separated to more dense than these particles. During centrifugation, the particles or molecules in the tube move up or down to the level at which the density of the medium is equal to their own density; at this level, they are said to be *isopycnic* with (equally as dense as) the medium. Even under tremendous centrifugal force, a particle will not sediment through a gradient region denser than itself.

Probably the most commonly used material for making density gradients in equilibrium centrifugation is a water solution of cesium chloride (CsCl). The cesium ion (Cs^+) is so compact that it sediments slightly in the powerful fields created in modern ultracentrifuges. A gradient is thereby established, with more Cs^+ (and more Cl^-, which follows the Cs^+ to neutralize the charge) toward the bottom of the tube. In a typical ultracentrifuge run, the liquid will be about 0.02 g/mL denser at the bottom of the tube than at the top. Thus molecules that differ in density by even a fraction of 0.02 g/mL can easily be separated by this technique. The densities of protein, DNA, and RNA in a solution of CsCl are approximately 1.3, 1.6–1.7, and 1.75–1.85 g/mL, respectively, so these molecules are easily separated from one another. The densities given here are higher than those of the same macromolecules in cells, because ions in a CsCl solution bind to proteins and nucleic acids; the densities of all macromolecules without bound ions are almost equal: 1.25–1.3 g/mL. In a CsCl solution, the Cs^+ binds to DNA mainly at phosphate groups; it binds to RNA both at phosphates and the hydroxyl groups of riboses, thus increasing the density of RNA more than that of DNA. Proteins, which have much less average charge than nucleic acids do, bind less cesium.

Proteins or nucleic acids in which ^{13}C or ^{15}N is substituted for ^{12}C or ^{14}N in amino acids or nucleotides also can be separated from their normal counterparts. For example, since proteins are 14 percent nitrogen and ^{15}N is $^{15}/_{14}$ times as dense as ^{14}N (Table 6-2), a protein substituted completely with ^{15}N is about 1 percent denser than the normal protein—a sufficient difference to allow the normal protein to completely separate from the substituted one. Thus, when cells are grown in a medium containing heavy amino acids or nucleotide precursors, it is possible to physically separate molecules made by the cells before and after the addition of the heavy isotope (see Figure 12-1).

Table 6-2 Commonly used heavy isotopes and their natural (more abundant) counterparts*

Heavy isotope	Atomic mass	Natural isotope	Atomic mass
Deuterium (hydrogen 2)	2.01	Hydrogen 1	1.01
Carbon 13	13.01	Carbon 12	12.00
Nitrogen 15	15.00	Nitrogen 14	14.01
Oxygen 18	18.00	Oxygen 16	16.00

*The greater density of heavy isotopes is due to the presence of one or more additional neutrons in their nuclei. The extra neutrons do not affect the chemical bonding properties of the atoms but do affect their mass.

The Sedimentation Constant When a particle suspended in a medium is subjected to centrifugal force, it will move if its density d is greater than the density of the surrounding medium d_0. The speed of movement in a stationary medium is proportional to the gravitational acceleration g; in a centrifugal field, g is replaced by the centrifugal acceleration c, which is equal to $(2\pi\omega)^2 x$, where ω is the revolutions per unit of time and x is the distance of the particle from the axis of rotation.

As the particle moves, it encounters friction with the medium. As it accelerates, its velocity v increases, increasing friction. The frictional force ϕ is equal to fv, where f is a frictional coefficient related to the shape of the particle. For a spherical particle, $f = 6\pi\eta r$, where η is the viscosity of the medium and r is the radius of the particle. The velocity increases until the frictional force balances the centrifugal force P_c, after which time the particle continues to move at a uniform velocity. *Stokes's equations*

$$\tfrac{4}{3}\pi r^3(d - d_0)c = 6\pi\eta r v$$

$$v = 2cr^2\left(\frac{d - d_0}{9\eta}\right)$$

$$r = \sqrt{\frac{9v\eta}{2c(d - d_0)]}}$$

describe the motion of spherical particles in a fluid under ideal conditions (larger particles than solvent molecules; no interaction among particles; no disturbance due to convection, or heat transfer).

The sedimentation constant $v/c = s$ is characteristic for a given particle in a given medium at a given temperature. If r and x are expressed in cm, g or c in cm/sec^2, ω in r/sec, d in g/cm^3, and η in g/cm/sec, then the sedimentation constant s is calculated

$$s = \frac{v}{c} = \frac{\tfrac{4}{3}\pi r^3(d - d_0)}{6\pi\eta r} = \frac{m[1 - (d_0/d)]}{f}$$

where m is the mass (in g) of the spherical particle and s is expressed in sec or in *svedbergs* (S $= 10^{-13}$ sec, under standard conditions of sedimentation in water at 20°C, which standardize the friction of the sedimenting particles). Because the centrifugal force and the density and viscosity of the medium can all be measured, Stokes's equations, taken together, can be used to estimate the radius and mass of the spherical particle, if its density and sedimentation velocity are measured in a centrifuge. The s values for a representative set of biologically important particles are given in Table 6-3.

Table 6-3 Sedimentation constants and molecular weights for some molecules and other particles

Molecule or particle	Sedimentation constant (S)*	Molecular weight $\times 10^{-3}$
PROTEINS		
Cytochrome *c*	1.7	13.4
Myoglobin	2.0	16.9
Hemoglobin ($\alpha_2\beta_2$)	4.5	64.5
Fibrinogen	7.6	340
RNA		
Transfer RNA (average)	4.0	25–27
Ribosomal RNA:		
E. coli, small	16	550
E. coli, large	23	1100
Human, small	18	660
Human, large	28	1700
PARTICLES		
Ribosome (human)	80	—
Poliomyelitis virus	150	—
Bacterium	5000	—

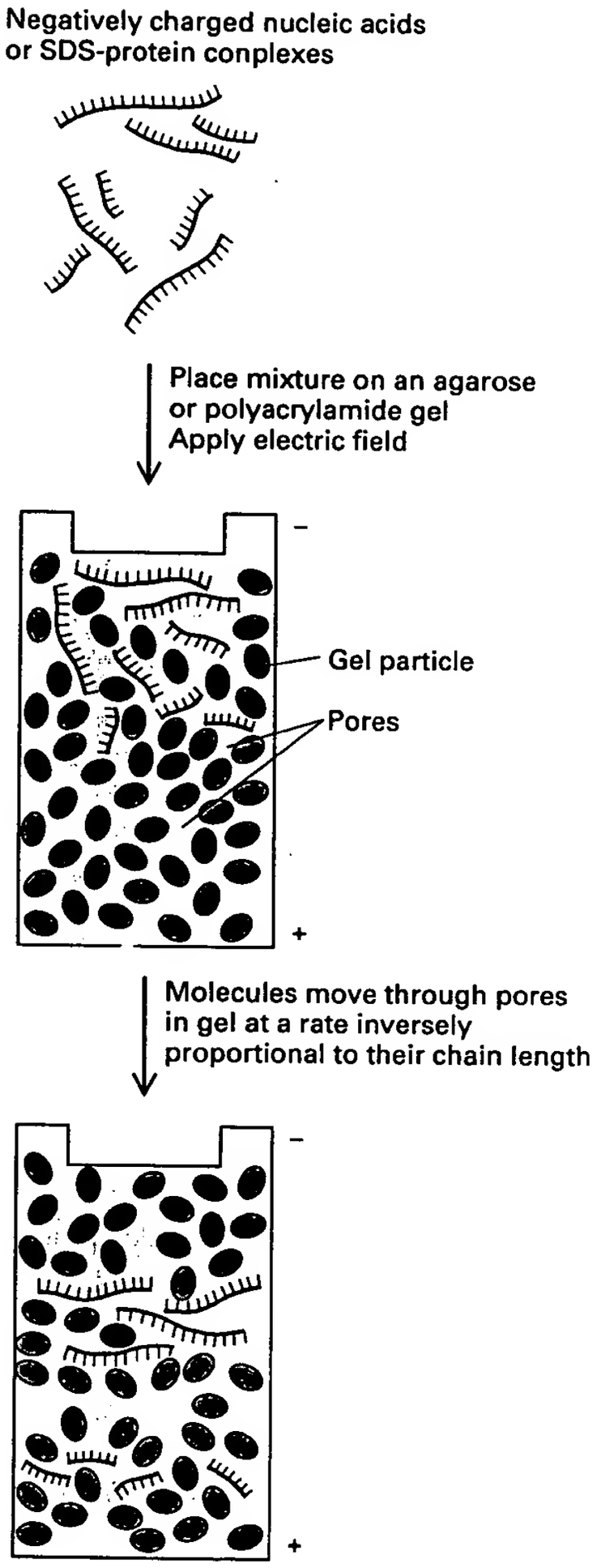
*S $= 10^{-13}$ sec.

Electrophoresis Separates Molecules According to Their Charge-Mass Ratio

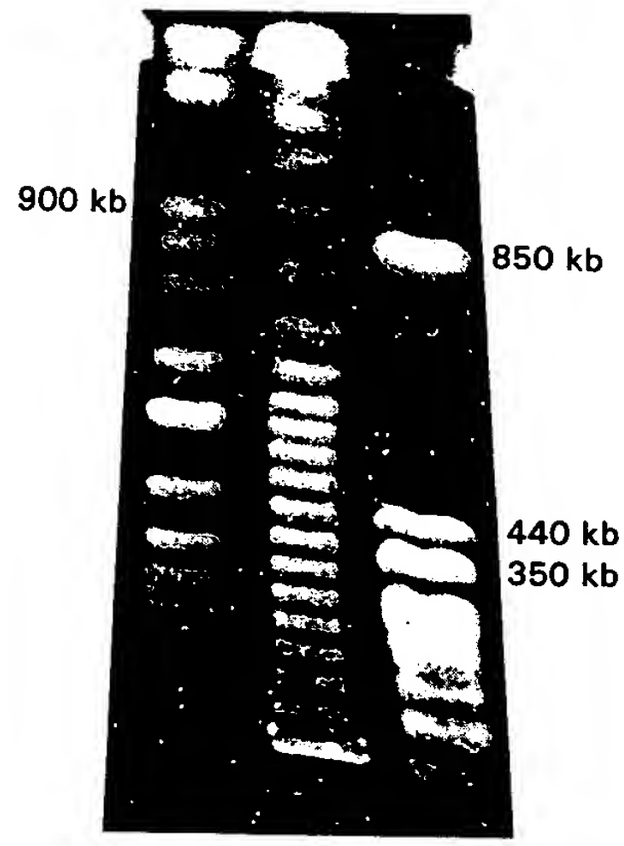
Molecules in a mixture can be separated, or *resolved,* according to size by *electrophoresis,* a technique dependent on the fact that dissolved molecules in an electric field move at a speed determined by their charge-mass ratio. For example, if two molecules have the same mass and shape, the one with the greater charge will move faster toward an electrode. Many successful variations of electrophoresis are in general use; the separation of small molecules, such as amino acids and nucleotides, is one example. A small drop of sample is deposited on a strip of filter paper or other porous substrate, which is then soaked with a conducting solution. When an electric field is applied at the ends of the strip, small molecules dissolve in the conducting solution and move along the strip at a rate corresponding to their charge.

Nucleic acids in solution generally have a negative charge because their phosphate groups are ionized; thus they migrate toward a positive electrode. However, nucleic acid molecules consisting of long chains have almost identical charge-mass ratios, whatever their length, because each residue contributes about the same charge and mass. Also, many proteins that differ in shape and mass have almost equal charge-mass ratios. Therefore, if the electrophoresis of nucleic acids and proteins were simply carried out in solution, little or no separation of molecules of varying lengths would occur.

Despite these difficulties, electrophoretic separation according to chain length has become amazingly reliable. Molecules are now most commonly subjected to electrophoresis in a *gel* (a semisolid suspension in water), rather than a liquid solution. The size of the pores in such gels limits the rate at which molecules can move through them. Nucleic acids with identical charge-mass ratios separate according to length, with the longer ones moving



▲ **Figure 6-7** Gel electrophoresis is carried out by pouring a liquid containing either melted agarose or chemically treated polyacrylamide into a cylinder (for a round gel) or between two flat, parallel glass plates 1–2 mm apart. As the gel solidifies, it forms interconnected pores, or channels, whose size depends on the concentration of agarose or polyacrylamide. The substances to be separated are then layered on top of the gel (or at one edge of it if it lies between two plates), and an electric current is passed through the gel. In usual laboratory practice, the migration of RNA or DNA depends on the charges on the phosphates: at neutral pH, a nucleic acid bears one negative charge per phosphate. Proteins can be separated by binding sodium dodecyl sulfate (SDS) to their amino acid residues, which contributes approximately one negative charge per residue. If all the particles have about the same charge-mass ratio, they move through the gel at a rate inversely proportional to their chain length.



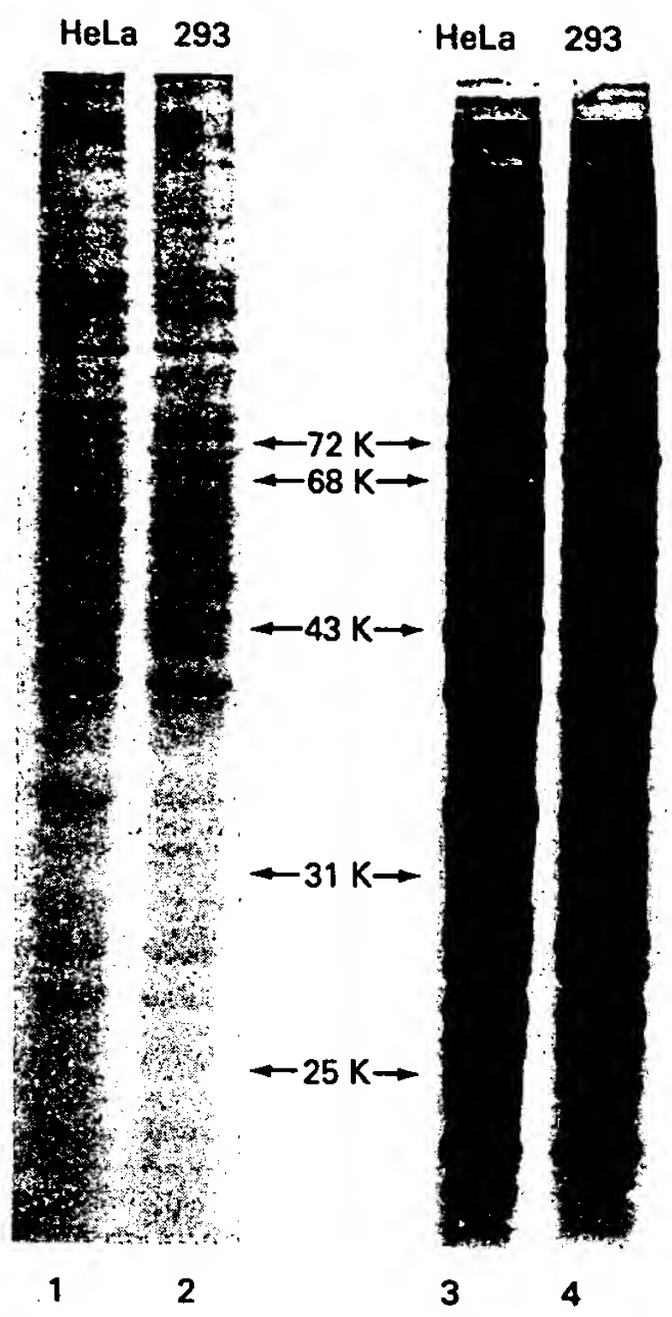
▲ **Figure 6-8** Pulse-field gel electrophoretic separation of large DNA molecules. In this technique, DNA molecules are moved first in one direction by application of an electric field. As they move, the molecules stretch out lengthwise in the direction of the field. The current is then stopped for a short time, and the molecules begin to "relax" into random coils; the time required for relaxation depends on the length of a molecule. The electric field can then be reapplied at 90° to the first direction or opposite to the first direction. Longer molecules relax more slowly than shorter ones, and so take longer to start moving in the new direction. Repeated alternation of field direction thus separates the molecules between the two directions and makes it possible to separate giant DNA molecules of 10^6 base pairs and more. The "ladder" in the lane 2 shows concatemers (linked units) of bacteriophage λ DNA in which each unit is 48.5 kb long (the band on the bottom is a single unit). Comparison with this ladder allows calculation of the length of other long DNA fragments. Lane 1 shows individual DNA molecules that each represent one chromosome from *Saccharomyces cerevisiae;* lane 3 shows a restriction digest with enzyme *Not*I that was used to map the *E. coli* chromosome (see Figure 5–9). [See C. L. Smith et al., 1987, *Science* 236:1448; C. L. Smith et al., 1987, *Nuc. Acids Res.* 15:4481.] *Photograph courtesy of C. L. Smith.*

more slowly (Figure 6-7). Even very long nucleic acids (chains containing 10,000–20,000 residues) that differ in length by only a few percentage points can be separated. In mixtures containing chains of 500 nucleotides or less, *each chain length can be resolved,* which has made DNA sequencing possible.

By employing the new technique of *pulse-field gel electrophoresis,* different-sized double-stranded DNAs in the range of 1–10 million base pairs (bp), or 1–10 megabases (Mb), can now be separated (Figure 6-8). Electrophoretic migration is begun in one direction; then the current is briefly stopped and reapplied at a 90° angle or in the

opposite direction. These long molecules tend to align along the electric field when the current is on and to relax when it is off. Relaxation time is affected by pores in the gel: longer molecules take longer to relax, and so respond more slowly as the current is switched, than shorter ones do, allowing the chains to be separated. This technique is very important for purifying long DNA molecules. It is required for the analysis of cellular chromosomes, which range from the smallest yeast chromosomes (about 5×10^5 bp) to the largest animal and plant chromosomes (2 or 3×10^8 bp).

Protein chains also can be separated according to length. Before and during electrophoresis, the proteins are continuously exposed to the detergent SDS (sodium dodecylsulfate, a common commercial cleaning agent found in toothpaste). Approximately one molecule of detergent binds to each amino acid. At neutral pH, the detergent is negatively charged; the adjacent negatively charged SDS molecules repel one another, forcing the proteins with bound detergent into rodlike shapes endowed with similar charge-mass ratios. Proteins in this state are said to be *denatured*. As with nucleic acids, chain length (which reflects mass) is the determinant for the separation of proteins by electrophoresis through polyacrylamide gels (Figure 6-9). Even chains that differ in molecular weight by less than 1 percent can be separated.



▲ **Figure 6-9** Resolution of proteins by one-dimensional gel electrophoresis. The proteins of two human cell lines—HeLa, a human cervical cancer cell, and 293, a virus-transformed embryonic fibroblast—were dissolved in SDS and subjected to electrophoresis. The newly made proteins are visible in lanes 1 and 2 by autoradiography, because the cells were labeled with [^{35}S]methionine, and in lanes 3 and 4 by the dye Coomassie blue, which stains all proteins. The designations 72 K, 68 K, etc., indicate the positions of marker proteins (proteins of known sizes) with molecular weights of 72,000, 68,000, etc. The major proteins in these two cell types are obviously quite similar. *Photographs courtesy of J. R. Nevins and C. Lawrence.*

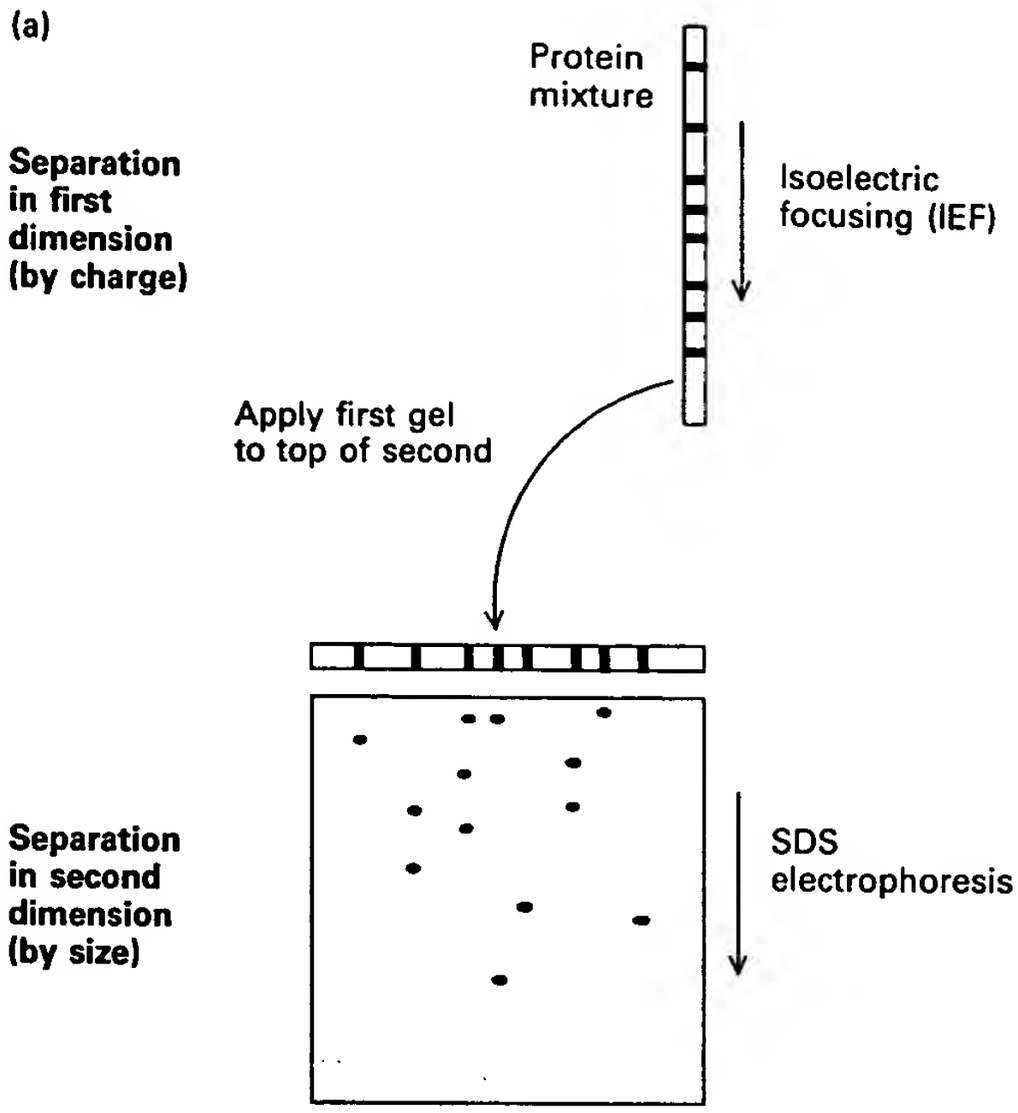
Gel Electrophoresis Can Separate Most Proteins in a Cell

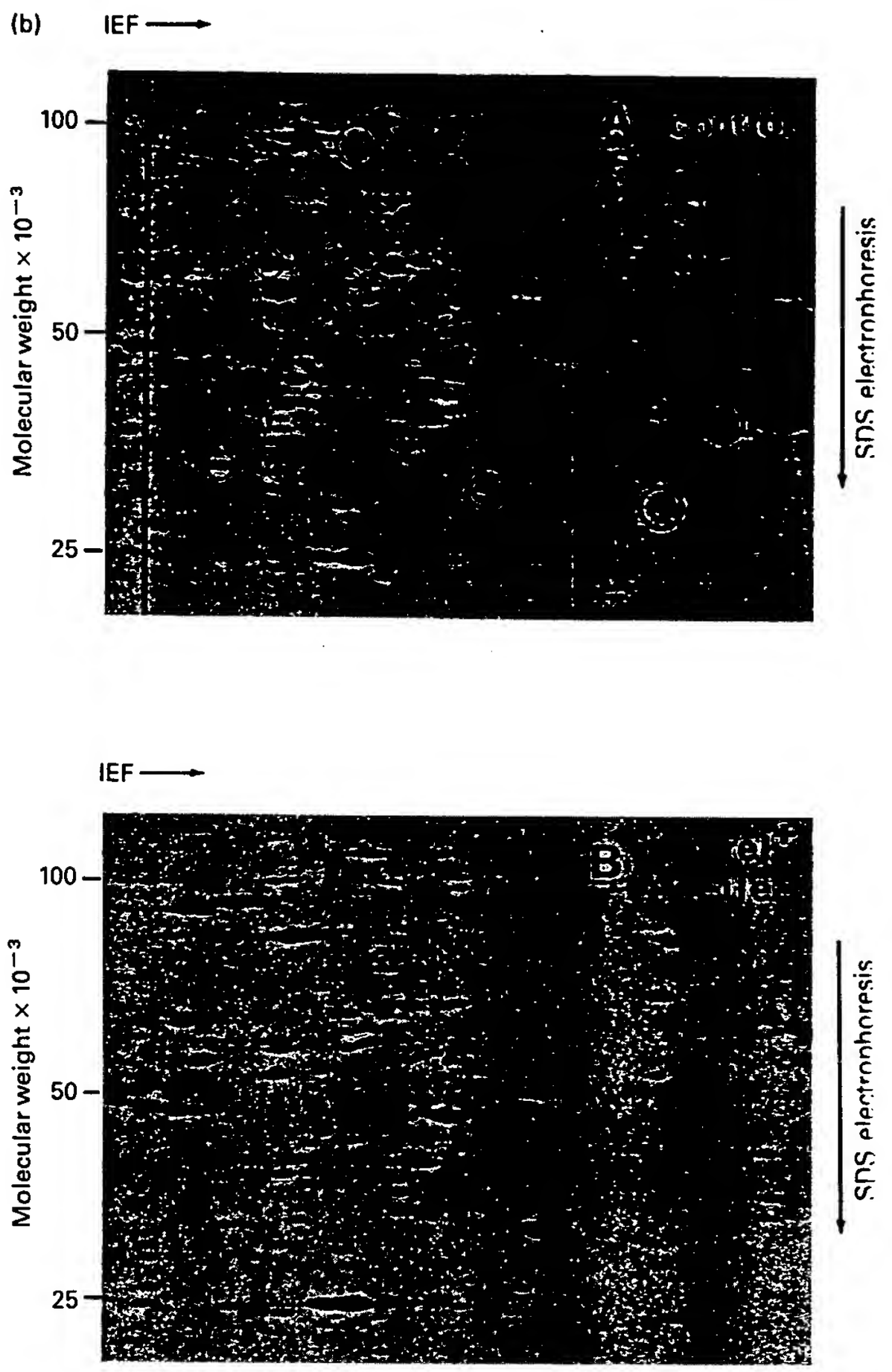
The traditional biochemical approach first to enzyme detection and ultimately to detailed enzyme chemistry is to detect enzymatic activity in a sample from a natural source and isolate the proteins that catalyze the activity. Biochemical methods of separating pure proteins from natural mixtures rely on differences in sedimentation rate or in charge change related to varying salt concentrations or pH. This causes the protein to bind differentially to various substances (e.g., cellulose products) and makes chromatography possible.

However, many experiments in molecular biology are designed to enumerate the polypeptides formed in a certain cell at a certain time, rather than to detect active enzymes or determine their concentrations. Sometimes just the presence of a given protein is to be detected anywhere within the cell, without purifying the protein. Or it may be important to compare the synthesis rate of a protein or a set of proteins with that of all other proteins in the cell, again without isolating any particular protein. Gel electrophoresis can often accomplish these aims.

Two-dimensional Gels Electrophoresis of all cellular proteins in one direction through a column or a thin rectangular SDS gel reveals only the major proteins. If these proteins are of interest or if a cell is producing large amounts of specific proteins (as occurs during viral infection), then this one-dimensional analysis may suffice.

Resolution of virtually all proteins in the cell can be accomplished in a two-dimensional gel, which separates the proteins in a sample first by charge and then by size (Figure 6-10). Separation by charge is carried out by *isoelectric focusing* (IEF). A protein that has not been denatured with SDS has a characteristic overall charge on its surface, which varies with pH. When placed in a gradient of pH and subjected to an electric field, a protein will migrate to the pH at which its overall surface charge is neutral and remain at this *isoelectric point*. Proteins separated in a gel of this type can, while still in the gel, be layered on top of another gel soaked with SDS; thus the proteins can be separated by electrophoresis in a second





▲ **Figure 6-10** (a) The preparation of two-dimensional protein gels by isoelectric focusing (IEF) followed by electrophoresis. (b) Labeled proteins can be detected by autoradiography. Each spot represents a single polypeptide. The spots are elongated horizontally because the average charge on a protein molecule varies somewhat during IEF. These patterns are reproducible, so that changes in individual proteins can be detected. The proteins are from cells growing on a normal medium supplemented with isoleucine *(top)* and from cells placed for a brief period in a medium that lacks isoleucine *(bottom)*. Certain spots *(circles)* are absent in the bottom photograph or are much fainter than in the top photograph; these differences represent changes in the synthesis pattern of cell proteins in response to amino acid starvation. *From P. H. O'Farrell, 1978,* Cell *14:545. Copyright M.I.T.*

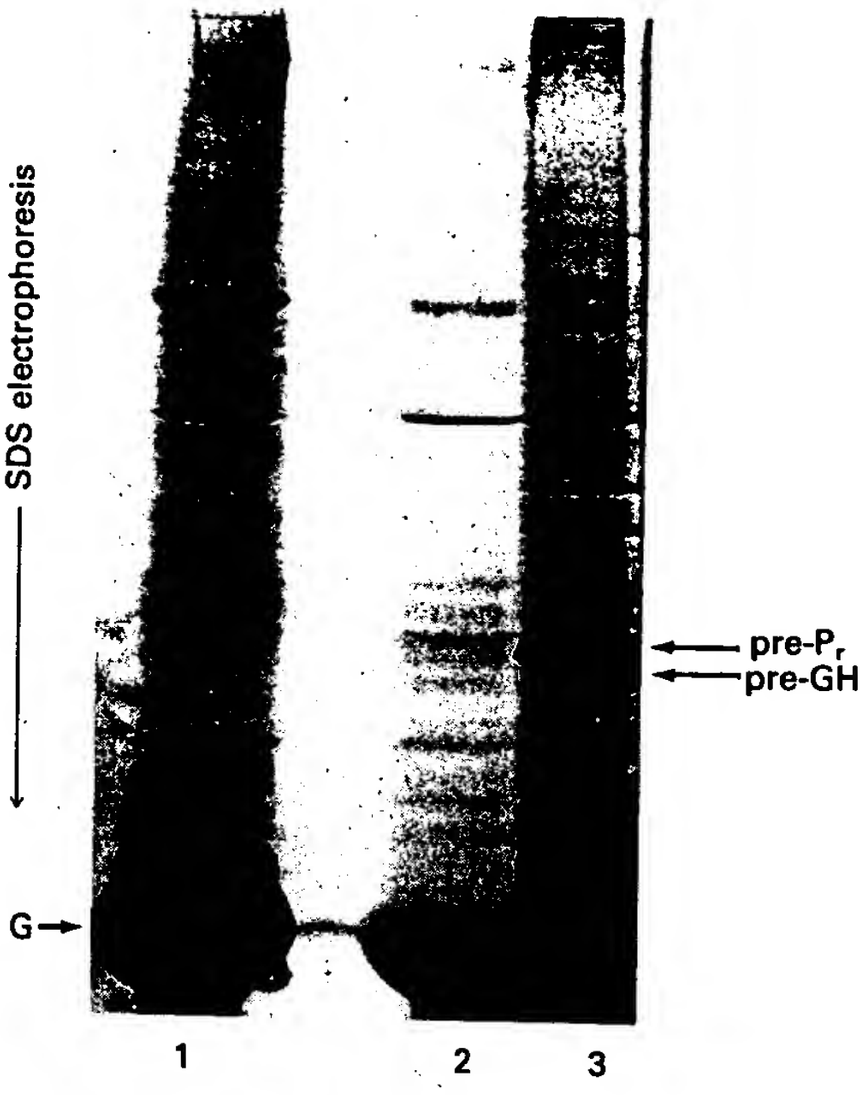
dimension on the basis of size. As many as several thousand different protein chains—virtually the total protein content of a cell—can be detected and separated by this technique. Two-dimensional gels are very useful in studying the expression of various genes in differentiated cells.

There are two widely used methods of detecting proteins in gels:

1. The total amount of each type of protein in a sample can be estimated with gel electrophoresis by staining the gels with a dye that binds approximately equally to all proteins. The intensities of the spots of dye indicate the comparative quantities of proteins of different lengths.
2. Gel electrophoresis provides a way of detecting the synthesis of any particular protein without isolating it. If whole cells are briefly labeled with radioisotopes before they are analyzed, each newly synthesized chain can be detected in the gel by autoradiography. However, because new proteins may be secreted from the cell or may be subject to different rates of metabolic turnover, the concentration of a labeled protein in a cell may not accurately reflect its rate of synthesis.

In Vitro Protein Synthesis and Gel Electrophoresis Provide an Assay for Messenger RNA

Two general approaches are used to determine what proteins a cell can make. In one method, the contents of whole, labeled cells are examined for newly synthesized proteins (see Figures 6-9, lanes 1 and 2, and 6-10). In the other, mRNA is extracted from the cells and translated in



▲ **Figure 6-11** The translation of mRNA by mixtures of ribosomes, tRNAs, and protein synthesis factors extracted from reticulocytes. Here, the total protein produced by such reactions has been separated by electrophoresis and is visible by autoradiography because [^{35}S]methionine was added to the extract. Proteins synthesized from mRNAs in an untreated reticulocyte extract are shown in lane 1; note the large amount of globin synthesis (G). In lane 2, a bacterial nuclease (from *Micrococcus aureus*) has greatly reduced the amount of synthesis. In lane 3, the nuclease has been chemically inactivated, and mRNAs from rat pituitary cells have been added. Several prominent pituitary-specific proteins are visible, including two hormone precursors: one of prolactin (pre-P$_r$, 236 amino acids long), and one of growth hormone (pre-GH, 212 amino acids long). [See H. R. B. Pelham and R. J. Jackson, 1976, *Eur. J. Biochem.* 67:247.] *Photographs courtesy of D. Anderson.*

the presence of labeled amino acids by cell-free protein-synthesizing systems (Figure 6-11). Both approaches are actually assays for functional mRNAs. In either case, the products can be separated and identified by gel electrophoresis.

Different cell extracts can be used to label proteins in vitro (assay for active mRNAs). Bacterial cell extracts that can translate homopolymers were first widely used to break the genetic code and to examine bacterial and bacteriophage proteins; now extracts of eukaryotic cells are also commonly used. Two of the most popular cell-free systems are extracts of reticulocytes and of wheat germ, the embryo plant in a fertile wheat seed. Both are prepared by treating the cells first with a nuclease that destroys endogenous mRNA (mRNA from the source cells) and then with a chemical that blocks the nuclease so subsequently added mRNA is not destroyed. After this treatment, very little protein synthesis by endogenous mRNA occurs (see Figure 6-11, lane 2). Thus the added mRNA is responsible for almost all protein synthesis, and the products of the added mRNA can be easily detected.

Examining the Sequences of Nucleic Acids and Proteins

The first biopolymer to be sequenced was a protein, and this discovery has great historical importance. Before Fred Sanger reported the sequence of human insulin in 1953, some biochemists were not convinced that proteins had specific sequences from end to end. The single unique sequence found in insulin implied a highly precise ordering mechanism during protein synthesis. Since that time, the coding of protein sequence by nucleic acid sequence has been made clear. Recently, it has become much easier to obtain long nucleic acid sequences than long protein sequences and thus, with the aid of the genetic code, to *deduce* the sequence of many proteins rather than actually determine them directly.

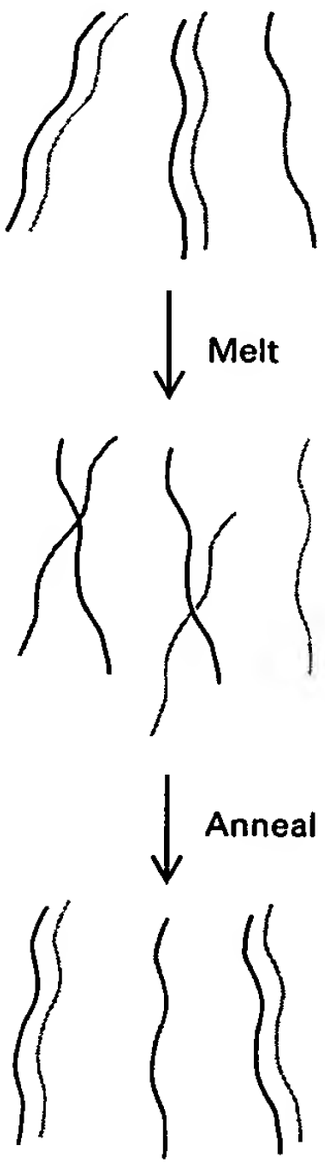
Because the functions of nucleic acids and proteins depend on the linear sequences of their monomers, research in molecular biology relies heavily on techniques that reveal and compare sequences. However, the sequence information required in experiments varies considerably in extent and type. In the simplest case, only an estimate of the degree of similarity, or *sequence relatedness,* between two samples of nucleic acid or protein is required. Often, it is necessary simply to determine whether a particular sequence is present in a given mixture of nucleic acids or proteins. Once the presence of a certain sequence in a mixture of sequences is established, a variety of other questions arise. What is the concentration or amount of the specific sequence? Where within a DNA, RNA, or protein molecule is the sequence of interest located? And finally, what is the precise nucleotide or amino acid sequence for the entire molecule? A variety of techniques are used to address these questions; each applies better to some questions than to others.

Molecular Hybridization of Two Nucleic Acid Strands Can Be Detected in Several Ways

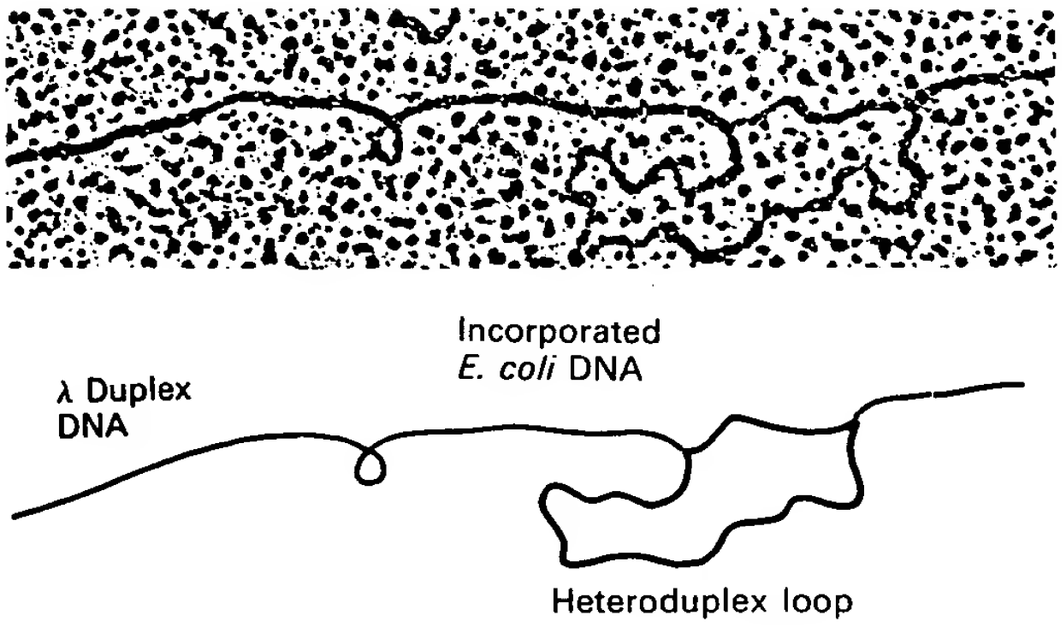
Under the conditions of temperature and ion concentration found in cells, DNA is maintained as a duplex (two-stranded) structure by the many hydrogen bonds of the A-T and G-C base pairs. The duplexes can be *melted* (*denatured* into single strands) by heating them (usually in a dilute salt solution of, for example, 0.01*M* NaCl) or by raising the pH above 11. If the temperature is lowered and the ion concentration in the solution is raised, or if the pH is lowered, the single strands will *anneal,* or reassociate, to reconstitute duplexes (if their concentration in

solution is great enough). In a mixture of nucleic acids, only complementary strands reassociate; the extent of their reassociation is virtually unaffected by the presence of noncomplementary strands. Such *molecular hybridization* can take place between complementary strands of either DNA or RNA or between an RNA strand and a DNA strand (Figure 6-12).

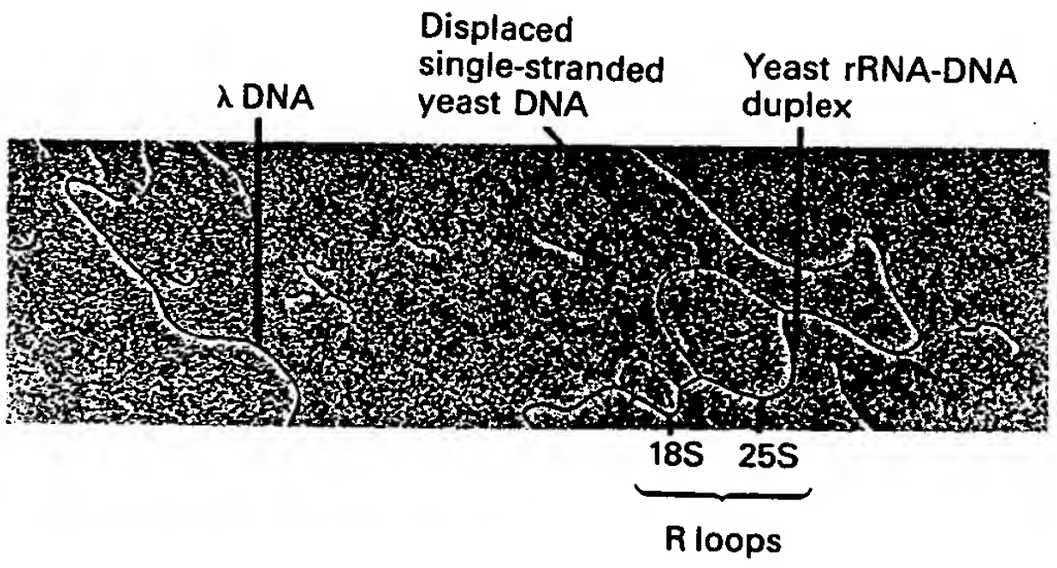
Visualization of Hybrids Electron microscopic examination of molecular hybrids conveniently reveals the sequence relatedness of two nucleic acid samples. If two melted nucleic acid samples that are complementary over only part of their length are allowed to hybridize, a *heteroduplex* results (Figure 6-13); complementary (duplex) and noncomplementary (single-stranded) regions can be distinguished in such preparations. This technique can be used not only to compare DNA strands but also to locate DNA sites complementary to RNA molecules. By this latter procedure, it is possible to distinguish and locate the regions of DNA that are transcribed into RNA. Regions of RNA-DNA hybridization create loops (called *R loops*) in the nucleic acid molecules, where the RNA sequence has base-paired with one DNA strand and displaced the other DNA strand (Figure 6-14). The fact that 1 μm of



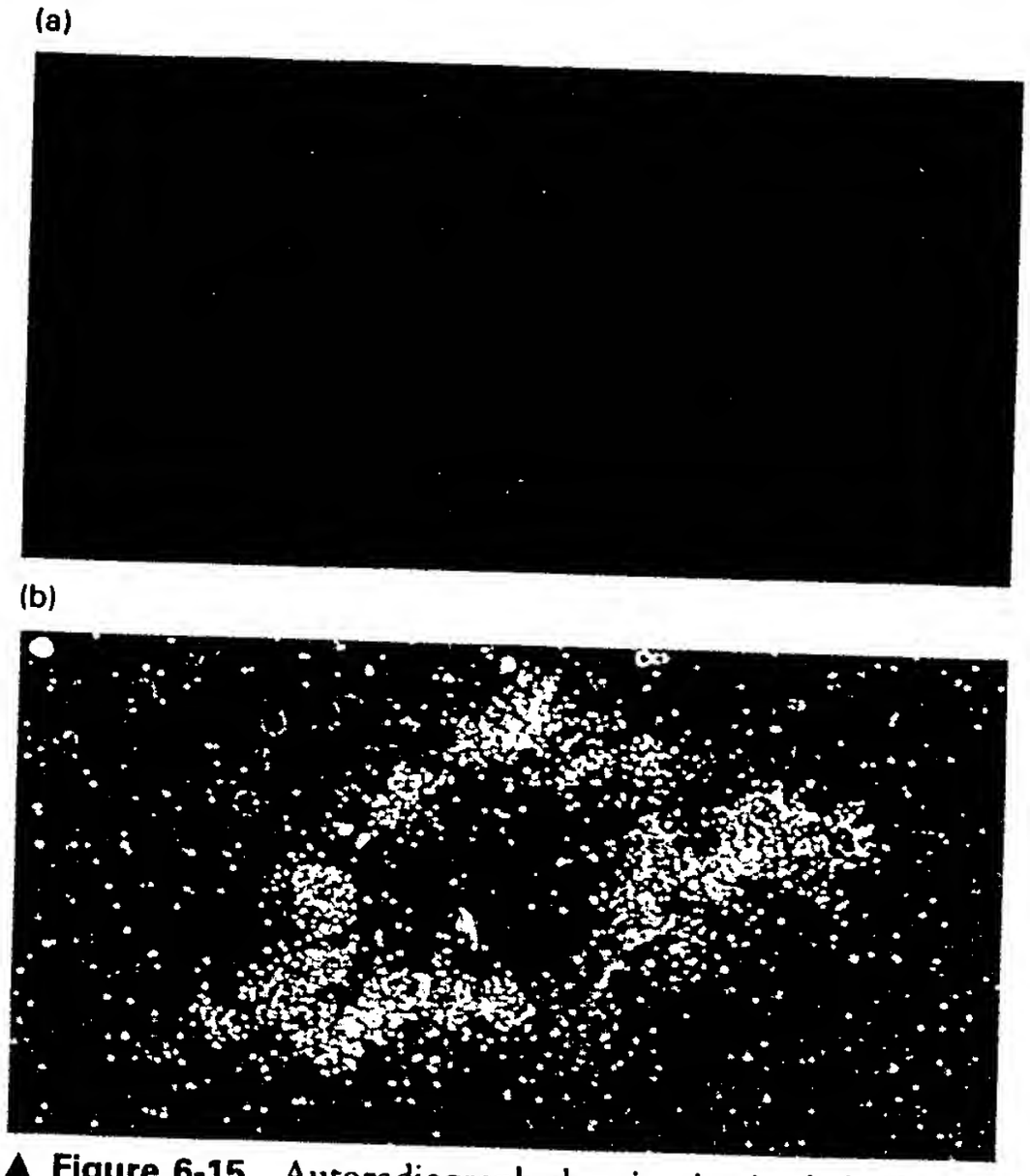
▲ **Figure 6-12** Molecular hybridization: reassociation of the complementary strands of a nucleic acid. Under conditions of high pH or temperature, the duplexes in a solution of nucleic acids melt, or separate into single strands. With an appropriate change in conditions, complementary strands reassociate. The presence of noncomplementary chains does not affect the reassociation rate of complementary chains.



▲ **Figure 6-13** Electron micrograph *(top)* of a DNA heteroduplex. DNA molecules on a carbon grid can be distinguished as long threads when they are shadowed with heavy metals (here, platinum and palladium). This heteroduplex has formed from strands of two λ bacteriophages incorporating different but related sequences of *E. coli* DNA *(bottom)*. The λ strands form a double-stranded hybrid where the inserted *E. coli* sequences are complementary (red); the dissimilar inserted sequences remain unassociated, resulting in a heteroduplex loop of single-stranded DNA (blue). *From R. W. Davis and J. S. Parkinson, 1971,* J. Mol. Biol. *56:403.*



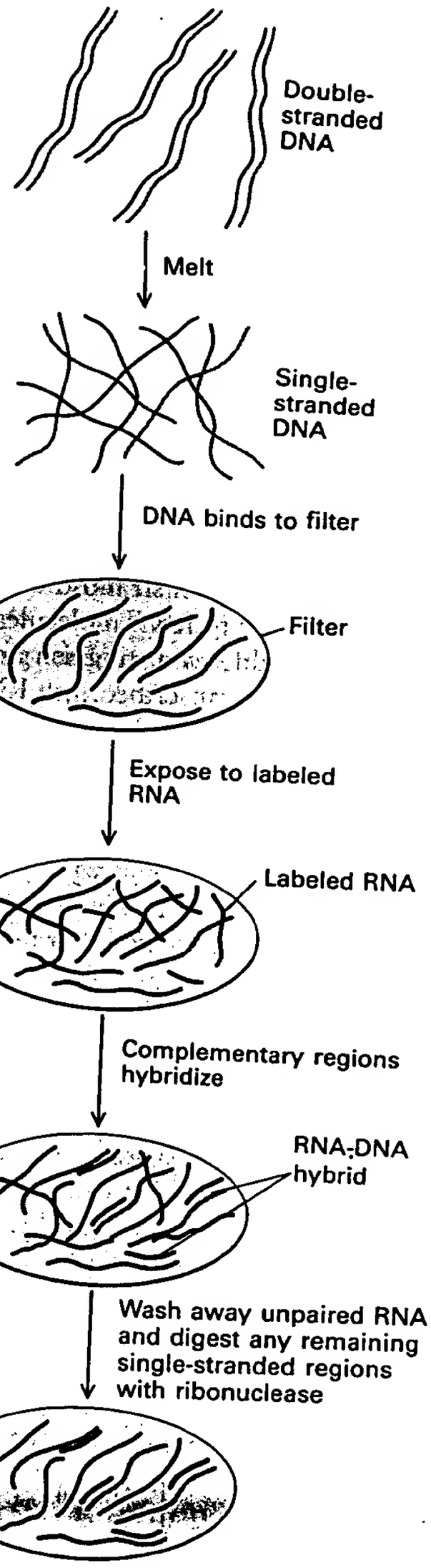
▲ **Figure 6-14** If double-stranded DNA is treated with a 50-percent solution of formamide at room temperature (or 25°C), some hydrogen bonds between the strands of the molecule break, weakening but not completely melting the duplex. If RNA that is complementary to one strand of the duplex DNA is then introduced, the RNA binds to its complementary site on one DNA strand, displacing the other DNA strand. This occurs because an RNA-DNA duplex is more stable than a DNA-DNA duplex. The hybrid duplex and the displaced stretch of single-stranded DNA are called an *R loop*. The two R loops that appear in this electron micrograph result from the hybridization of 18S and 25S ribosomal RNA from yeast with a region of bacteriophage λ DNA that contains an inserted stretch of yeast ribosomal genes. [See M. Thomas, R. L. White, and R. W. Davis, 1976, *Proc. Nat'l Acad. Sci. USA* 73:2294.] *Photograph courtesy of R. W. Davis.*



▲ **Figure 6-15** Autoradiograph showing in situ hybridization. A mouse liver section was exposed to labeled RNA complementary to glutamine synthetase mRNA. The label was allowed time to hybridize; then, unhybridized labeled RNA was washed away. (a) A light microscopic view of the autoradiograph showing cords of liver cells (hepatocytes) around a central vein. The barely visible dark grains around the central vein are in the first layer of hepatocytes. (b) The second view is a dark-field picture, which shows the grains (white dots) with much greater contrast. [See F. Kuo et al., 1988, *Mol. Cell Biol.* 8:4966.] *Photographs courtesy of F. Kuo.*

double-stranded nucleic acid contains about 3000 bases can be used to estimate the number of nucleotides in single- and double-stranded regions, and thus an accurate map of the transcribed section of DNA. The technique was crucial in proving splicing of mRNA in eukaryotes.

In Situ Hybridization Another use for molecular hybridization that has achieved great popularity is called *in situ hybridization*. Labeled RNA or DNA that is complementary to a specific mRNA is prepared. Cells or tissue slices are briefly exposed to heat or acid, which fixes the cell contents, including the mRNA, in place on a glass slide, the fixed cell or tissue is then exposed to the labeled complementary RNA for hybridization. Removal of unhybridized labeled RNA and coating the slide with a photographic emulsion is followed by autoradiography to reveal the presence and even the location of specific mRNA within individual cells (Figure 6-15).



▲ **Figure 6-16** The filter-binding assay for RNA-DNA (or DNA-DNA) hybridization is an extremely popular and flexible method of detecting complementary regions. Under the proper conditions of ion strength and temperature, filter-bound single-stranded DNA is exposed to a labeled RNA (or DNA) sample. Molecules or sequences complementary to the filter-bound DNA pair with it; unpaired labeled molecules can be removed. This technique allows as little as 1 part in 10^6 of specific RNA or DNA to be detected.

Hybrids on Filter-immobilized Nucleic Acid A common method of detecting hybrids between nucleic acid samples employs a single-stranded nucleic acid attached to a solid matrix. Nitrocellulose and treated nylon membranes are the most widely used matrices; it is not known why single-stranded DNA (or RNA) binds to these substrates, but this affinity is enormously useful. The radioactive RNA or DNA that is to be tested for sequences complementary to the bound nucleic acid is allowed to hybridize with it. After sufficient time, the unhybridized single strands unassociated with the bound nucleic acid are washed away. RNA hybridized to bound DNA is resistant to ribonucleases, whereas unpaired RNA regions are digested by these enzymes; thus any remaining single-stranded RNA regions are trimmed away in such experiments. The amount of hybrid formed can then be measured by the amount of bound radioactive label present (Figure 6-16). With the appropriate choice of filter-bound nucleic acid, one specific RNA (or DNA) sequence can be detected in a mixture of many different sequence types.

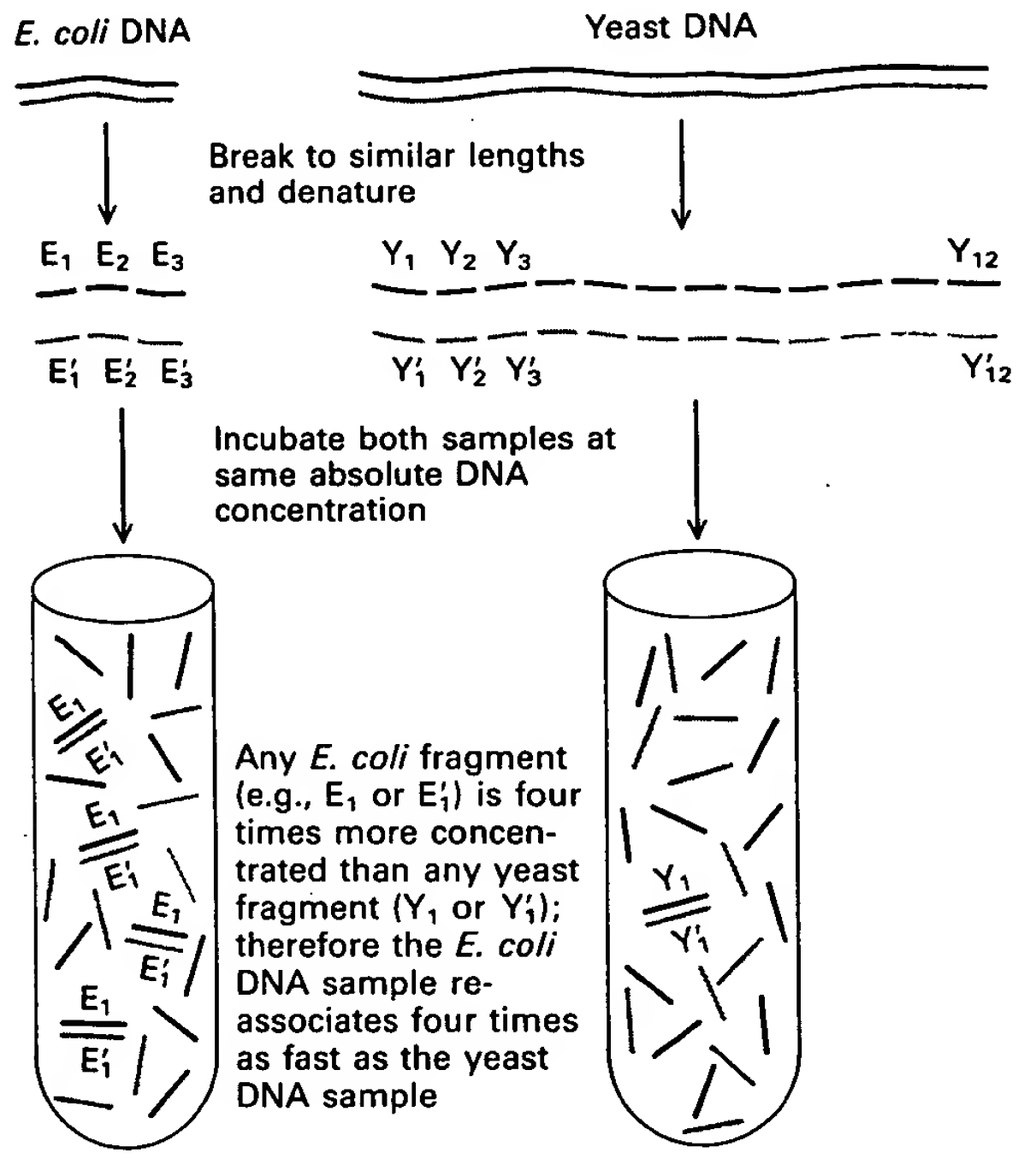
In the procedure called *DNA excess hybridization,* the total RNA from cells (a very complex mixture of sequences) is labeled and exposed to unlabeled purified specific DNA. If the DNA is present in excess (if there are more than enough copies to hybridize with all complementary segments of RNA), the amount of hybrid formed is proportional to the amount of RNA input. This allows an accurate measurement of the amount of the particular RNA in a mixed sample.

It is also possible to test for the presence of a particular RNA sequence and to quantify the amount of it in different samples by *competition hybridization.* A measured sample of a specific labeled RNA is exposed to just enough complementary DNA to completely hybridize with it; a sample of unlabeled RNA is then added. If the unlabeled RNA sample contains the same sequence as the labeled RNA, they "compete" for the DNA; increasing the ratio of unlabeled to labeled samples decreases the amount of labeled RNA hybridized. The extent to which this takes place is a measure of the amount of competing RNA in the unlabeled sample.

The Rate of Nucleic Acid Hybridization Can Be a Measure of Complexity The rate of hybridization between two complementary single-stranded nucleic acids in solution depends on the frequency with which complementary regions collide and *nucleate,* or start to form a duplex. This frequency, in turn, depends on the concentration of the two strands. If the DNA fragments of two different organisms—say, *Escherichia coli* and yeast—are incubated in amounts that yield the same *total* DNA concentration, the *complexity* of the DNA (the number of base pairs in the total genome) is about four times as great for yeast as for *E. coli.* A separated strand of *E. coli* DNA therefore encounters its correct partner four times as often as a strand of yeast DNA does, and *E. coli* DNA reassociates at a faster rate (Figure 6-17).

From the equation for determining the quantitative relation between reassociation rate and genome complexity (given in Figure 6-17), the reassociation rate of any DNA sample can be used to calculate the relative complexity of the source genome. Experimentally, the initial concentration of DNA C_0 and the time t are varied to measure the reassociation rate, so the resulting curves are often called C_0t *(cot)* curves.

If the DNA sequences of an organism are present once per haploid genome, the reassociation curve is uniform. If some sequences are repeated, these hybridize more rapidly. Reassociation measurements have been important both in comparisons between different types of organisms

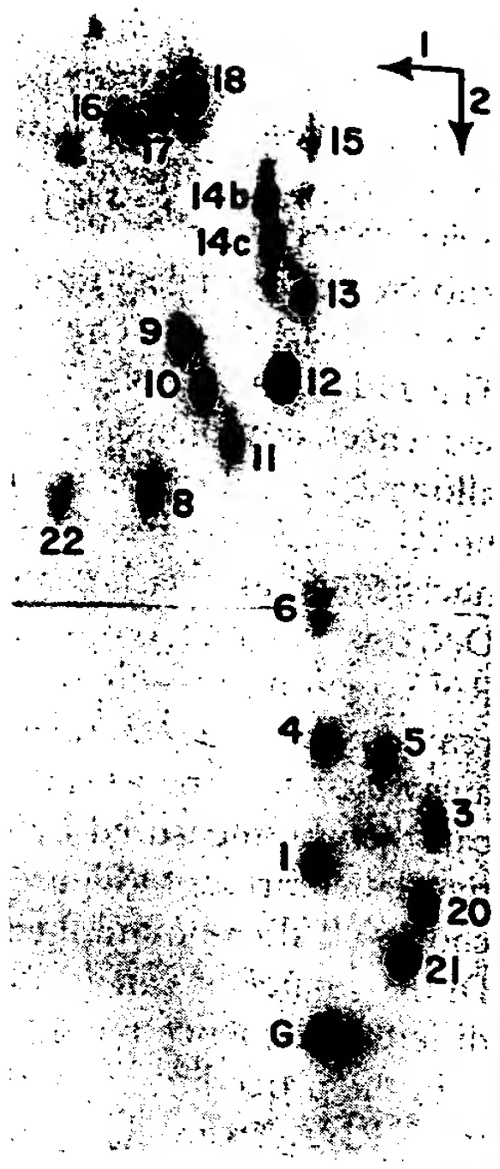


▲ **Figure 6-17** The complexity of DNA controls the rate of its reassociation. The relative hybridization rates within two samples of dissolved and melted genomic DNA depend on their relative complexity (the number of DNA base pairs in the genome of each organism), provided the samples are equal in absolute concentration (total nucleotide concentration). DNA is broken into pieces of 1000–2000 bases each, so size plays little role in the comparison. The equation for the reassociation rate is

$$\frac{C_t}{C_0} = \frac{1}{1 + KC_0t}$$

where C_0 and C_t are the molar concentrations of single strands at times 0 and t, respectively, and K is the rate constant for the particular type of DNA (this constant depends on the complexity of the DNA). [See R. J. Britten and D. E. Kohne, 1968, *Science* **161**:529.]

Spot number	Sequence	Spot number	Sequence	Spot number	Sequence
1	AG	10	UAG	16	UUAG
3	CCCG	11	UCG	17	UCUG
4	AAG	12	CCUG	18	AUCUCG
5	ACCG	13	CCUACG	20	CCG
6	AAAG	14b	AAUACCG	21	CG
8	CG	14c	AUCCAG	22	pG
9	AUG	15	CCACACCACCUG	G	Gp



▲ **Figure 6-18** A ribonuclease T1 fingerprint of 5S ribosomal RNA from oocytes of the frog *Xenopus laevis*. This enzyme cuts RNA on the 3′ side of all guanylate residues, NpGp ↓ Np(Np)$_n$Gp ↓ Np producing fragments that all contain one Gp (guanylate) at their 3′ ends. The digest is applied to treated paper (cellulose acetate), and a two-step separation is carried out: electrophoresis in one dimension *(arrow 1)*, followed by chromatography in the other *(arrow 2)*. If the starting sample is radioactive (^{32}P-labeled RNA is often used), the oligonucleotides can be identified by autoradiography. Spots of RNA can be cut out of the paper sheet and further analyzed biochemically to determine their sequences. No spots are numbered 2, 7, 14a, or 19; these numbers were given to oligonucleotides identified in another type of 5S rRNA. *From D. D. Brown, D. Carroll, and R. D. Brown, 1977,* Cell *12:1045. Copyright M.I.T.*

(prokaryotic versus eukaryotic; vertebrate versus invertebrate; and so on) and in studies of the degree of repetition of certain sequences within eukaryotic genomes. (Repetitious DNA sequences are discussed in detail in Chapter 10.) In a variation of the use of reassociation curves, a trace amount of radioactive pure DNA is added to unlabeled RNA from a cell of interest. Because the DNA is present in a tiny amount, the rate of RNA-DNA hybridization depends on the concentration of complementary RNA. From that rate, the amount of complementary RNA in a sample can be estimated. Data curves from such measurements are referred to as R_0t (*rot,* or RNA concentration) curves.

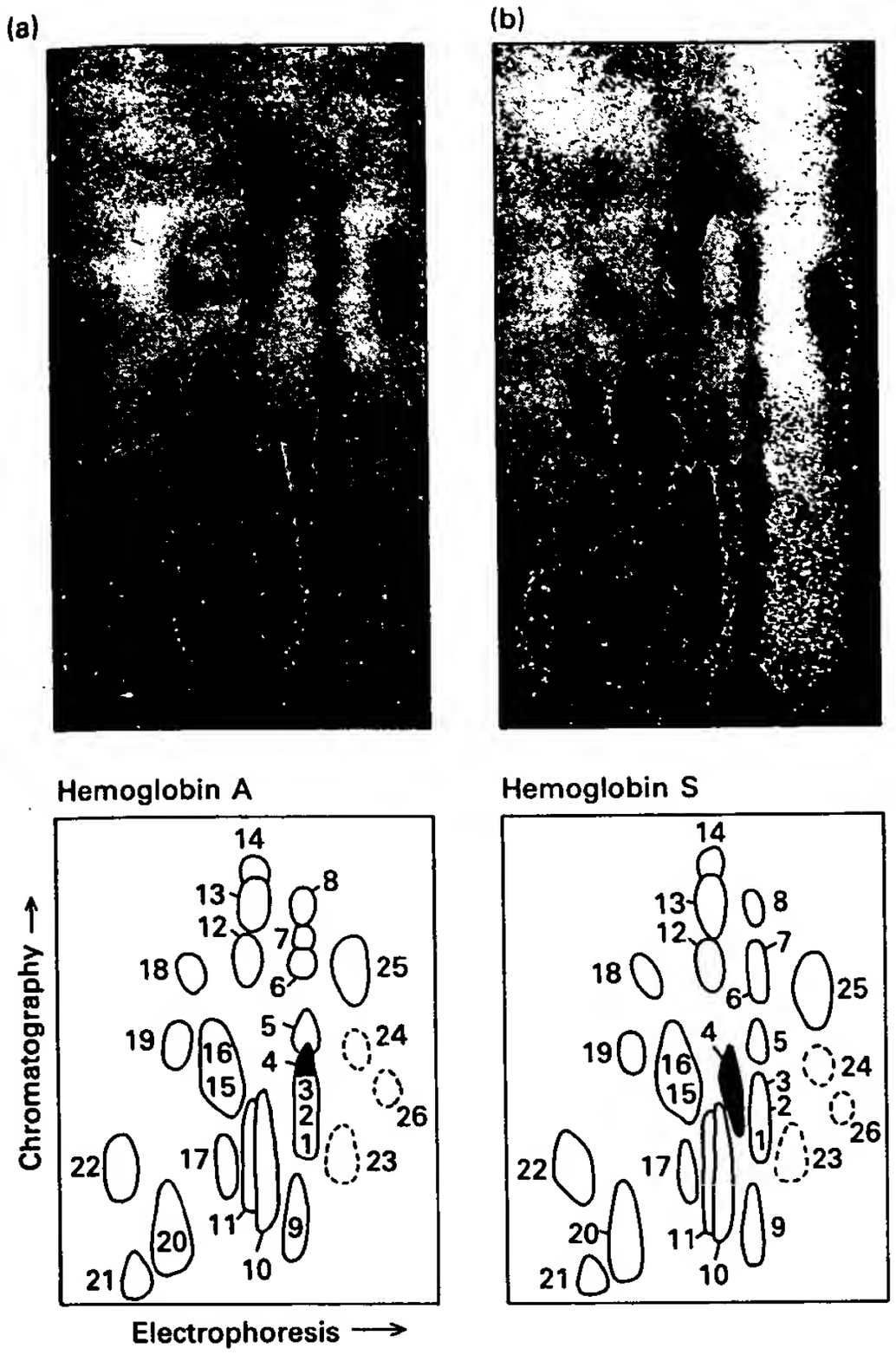
Fingerprinting (Partial Sequence Analysis) Allows Quick Comparisons of Macromolecules

The enzymatic fragmentation of proteins and nucleic acids at specific sites provides a means of recognizing particular macromolecules quickly. As we saw in Chapter 2, the enzyme trypsin digests protein chains, cleaving them on the carboxyl side of each lysine and each arginine residue to produce a specific set of peptides from any given pure protein. Likewise, the enzyme ribonuclease T1 cuts RNA at the 3′ side of each guanylate residue to produce specific fragments ending with a guanylate. The resulting oligonucleotides are fairly short: they normally contain 2–20 nucleotides, because consecutive guanylates are usually no more than 20 bases apart. Reliable separation and detection of different peptides from a pure protein or different oligonucleotides from a pure RNA sample can be accomplished by electrophoresis, chromatography, or both.

Because the oligonucleotides or peptides produced by a given enzyme from a pure RNA sample or protein are always the same, the pattern of separated fragments is always the same. The characteristic pattern of fragments from a primary sequence is called a *fingerprint* (Figures 6-18 and 6-19). *Fingerprinting,* or *partial sequence analysis,* allows the rapid comparison of two samples of RNA or protein when there is no need to determine the complete sequence of nucleotides or amino acids. The fingerprinting technique was first developed for proteins, which can be cleaved by both enzymes and chemical reactions. With this historic fingerprints of globin shown in Figure 6-19, Vernon Ingram demonstrated that people suffering from sickle-cell anemia, a genetic disease, have a valine substituted in one position in place of a glutamic acid in their β globin. This was the first mutant protein shown to be affected in function by a change in one amino acid residue.

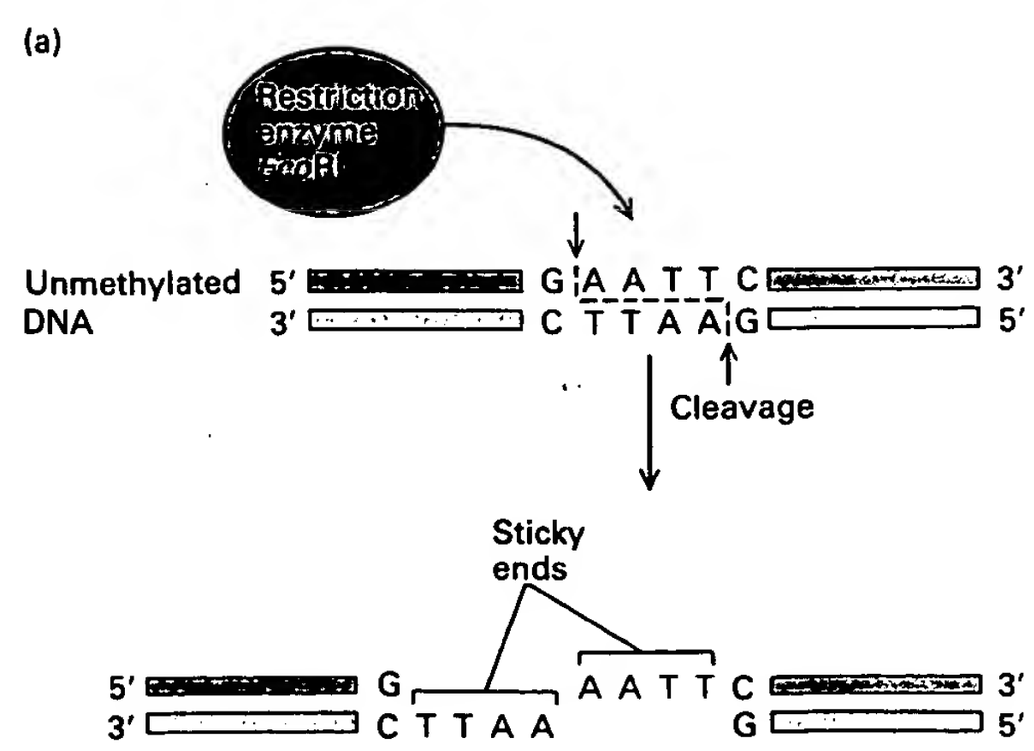
Restriction Enzymes Allow the Precise Mapping of Specific Sites in DNA

The most flexible, simple, and useful technique for partial sequence analysis of DNA was made possible by the discovery of bacterial restriction endonucleases, which recognize specific short oligonucleotides from four to eight residues long in DNA and then cleave the DNA at each site (Figure 6-20a). The word "restriction" refers to the



◀ **Figure 6-19** Fingerprints of (a) normal and (b) sickle-cell human β-globin. Proteolytic enzymes such as trypsin are used to break the peptide chain at known amino acid residues (trypsin cuts after each arginine and each lysine). The resulting set of specific fragments can then be separated by electrophoresis followed by chromatography. Individual peptide spots can be distinguished by spraying the chromatography paper with ninhydrin, a reagent that forms a purple product with free amino groups. (Spots 23, 24, and 26 show up poorly with ninhydrin.)

These fingerprints are identical, with one exception. Peptide 4 of the β chains of hemoglobin S (the hemoglobin of people with the sickle-cell disease) is found in a slightly different location than peptide 4 of normal hemoglobin A. Analysis of the two peptides has shown that hemoglobin S has a valine instead of a glutamic acid at residue 6 in the β chain; thus a single amino acid replacement is the cause of sickle-cell anemia. This represented the first demonstration that a random mutation in nature resulted in a single amino acid substitution. *From V. Ingram, 1958,* Biochim. Biophys. Acta *28:543.*



▼ **Figure 6-20** (a) *Eco*RI and many other restriction endonucleases cleave DNA so that the fragments have short complementary single-stranded segments at the ends. These "sticky ends" are important in recombinant DNA techniques because they readily pair with the ends of other cleavage fragments produced by the same restriction endonuclease. *Eco*RI recognizes the sequence shown here. (b) Most cells with restriction endonucleases also have corresponding *modification endonucleases. Eco*RI methylase, a modification endonuclease, catalyzes the methylation of two adenylates (shown in blue) in the recognition sequence; this prevents cleavage by *Eco*RI. Thus a cell making *Eco*RI endonuclease and methylase does not destroy its own DNA.

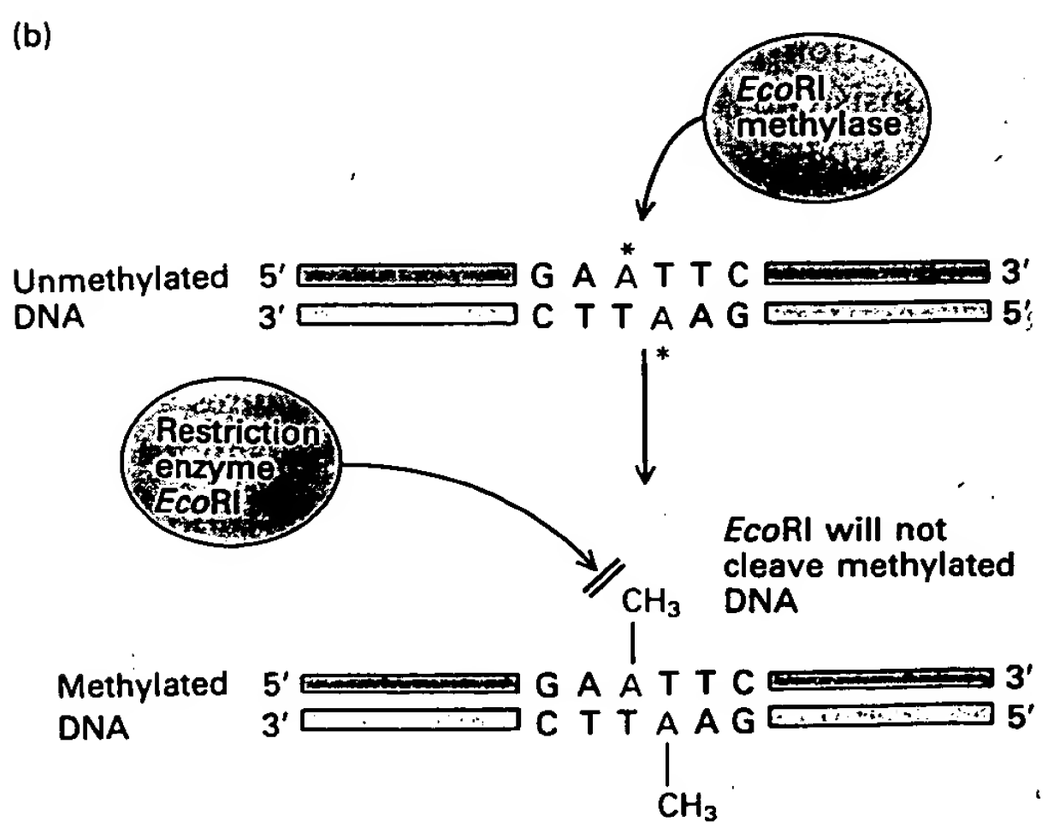


Table 6-4 Examples of the actions of restriction endonucleases

Source microorganism	Enzyme*	Recognition site (↓)†	Number of cuts (kb)‡			
			λ (50)	Ad2 (36)	SV40 (5.2)	pBR322 (4.3)
Arthrobacter luteus	*Alu*I	AG↓CT	143	158	34	14
Thermus aquaticus	*Taq*I	T↓CGA	121	50	1	13
Haemophilus parahaemolyticus	*Hph*I	GGTGA+5	168	99	4	18
Haemophilus gallinarum	*Hga*I	GACGC+8	102	87	0	12
Escherichia coli	*Eco*RI	G↓AATTC	5	5	1	1
Haemophilus influenzae	*Hin*dIII	A↓AGCTT	6	12	6	1
Nocardia otitiscaviarums	*Not*I	GC↓GGCCGC	0	7	0	0
Streptomyces fimbriatus	*Sfi*I	GGCCN$_4$↓NGGCC	0	3	1	0

*Enzymes are named with abbreviations of the bacterial strains from which they are isolated; the Roman numeral indicates the enzyme's priority of discovery in that strain (for example, *Alu*I was the first restriction enzyme to be isolated from *Arthrobacter luteus*).
†Recognition sequences are written 5′ → 3′ (only one strand is given). For example, G↓GATCC is an abbreviation for

(5′)G↓GATCC(3′)
(3′)CCTAGG↑(5′)

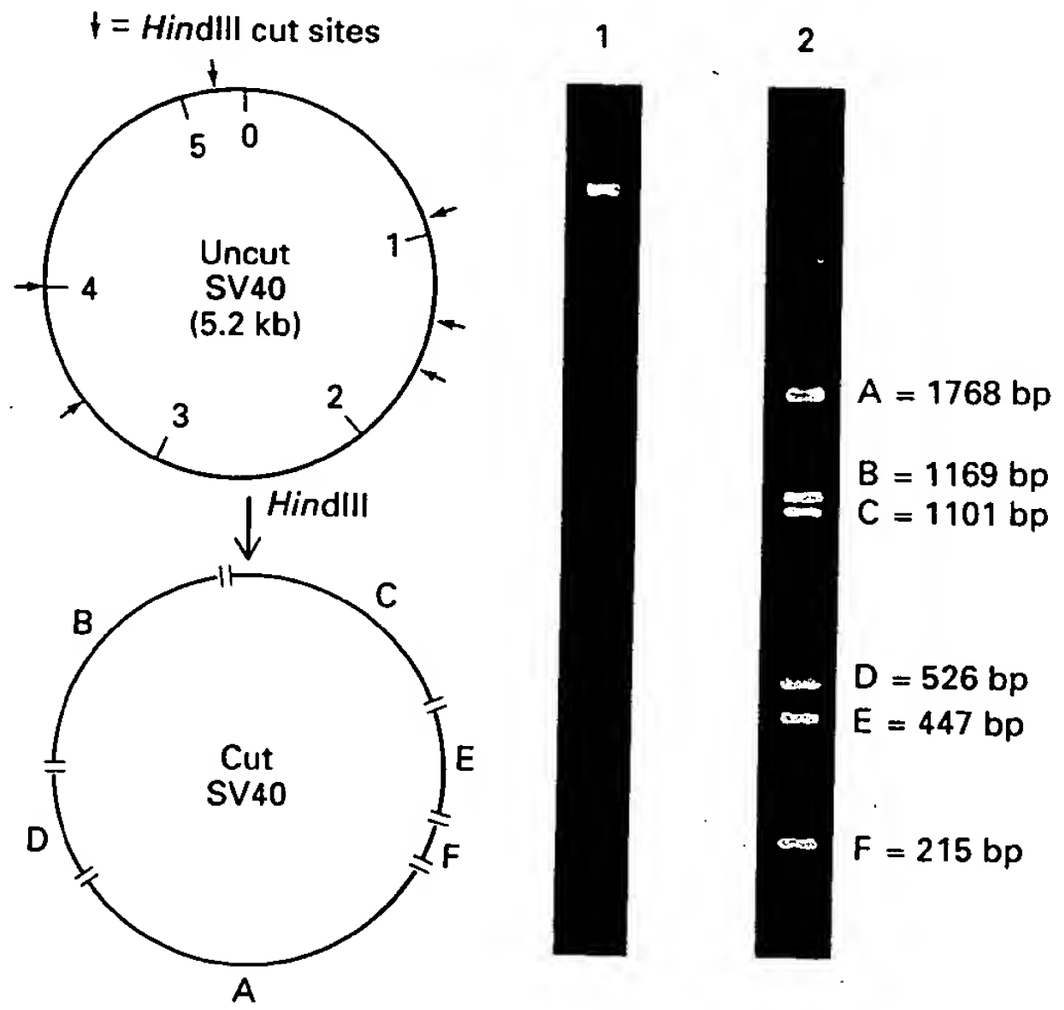
The cleavage site for *Hph*I and *Hga*I occurs five or eight bases away from the recognition sequence; N indicates any base.
‡These columns list the number of cleavage sites recognized by specific endonucleases on the DNA of bacteriophage λ (λ), adenovirus type 2 (Ad2), simian virus 40 (SV40), and an *E. coli* plasmid (pBR322). The sizes of the DNAs are given in kilobases (kb). Note that the actual number of cuts in these sequences deviates from the expected number in random sequences, which would be given by $L/4^n$, where n is the length of the site recognized by an enzyme and L is the length of the sequence.

SOURCE: R. J. Roberts, 1988, *Nuc. Acids Res.* **16**(supp):r271.

function of these enzymes in the bacteria of origin: a restriction endonuclease destroys (restricts) incoming foreign DNA (for example, bacteriophage DNA or DNA accidentally taken up during transformation) by cleaving it at these specific sites, called *restriction sites*.

Another enzyme protects a bacterium's own DNA from cleavage by modifying it at or near each potential cleavage site: a methylase adds a methyl group to one or two bases, usually within the restriction site. When a methyl group is present there, the restriction endonuclease is prevented from cutting the DNA (Figure 6-20b). Together with the restriction endonuclease, the methylating enzyme forms a *restriction-modification system* that protects the host DNA while it destroys foreign DNA.

A restriction endonuclease cuts a pure DNA sample into a consistently reproducible set of fragments that can be easily separated by gel electrophoresis (Figure 6-21). Several hundred restriction enzymes with different recognition sites are now available (see Table 6-4). If the order of nucleotides in DNA were random, the number of cuts expected would be larger for an enzyme that requires only a four-base site than for one that requires a longer site and larger for longer DNAs than for shorter ones. However, the sites for restriction endonucleases are not randomly distributed; by testing a series of enzymes that cut at different sites it is possible to cut a particular DNA many or only a few times. The most recently discovered eight-base cutters have proved to be especially useful for producing very large fragments, which then can be separated by pulse-field gel electrophoresis (see Figures 5-9



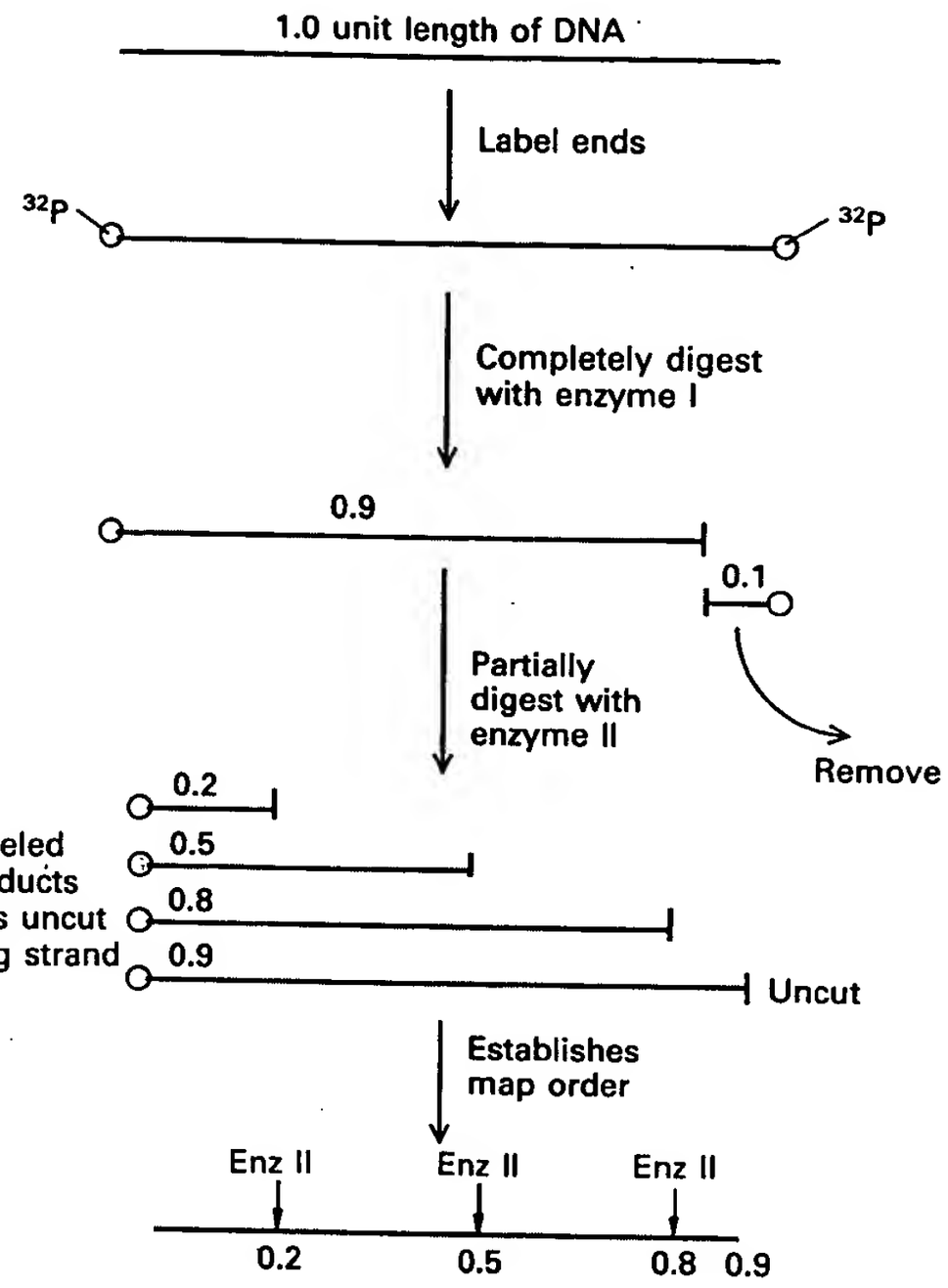
▲ **Figure 6-21** The DNA from SV40 virus can be purified and digested with the restriction endonuclease *Hin*dIII (from *Haemophilus influenzae*). The digest is then subjected to electrophoresis in a gel containing ethidium bromide, a molecule that binds to DNA and fluoresces when exposed to ultraviolet irradiation. Lane 1 represents the uncut DNA; lane 2, the digested DNA. *Hin*dIII cuts the SV40 molecule at six sites (↓), producing six fragments. By convention, the pieces of DNA released by a restriction endonuclease are labeled A–Z in order of decreasing size; the *Hin*dIII fragments of SV40 are therefore labeled A–F. *Photograph courtesy of D. Nathans.*

and 6-8). Fragments of 1–10 megabases (10^6–10^7 bp) are used to map the chromosomes of very large genomes, such as those of mouse and man.

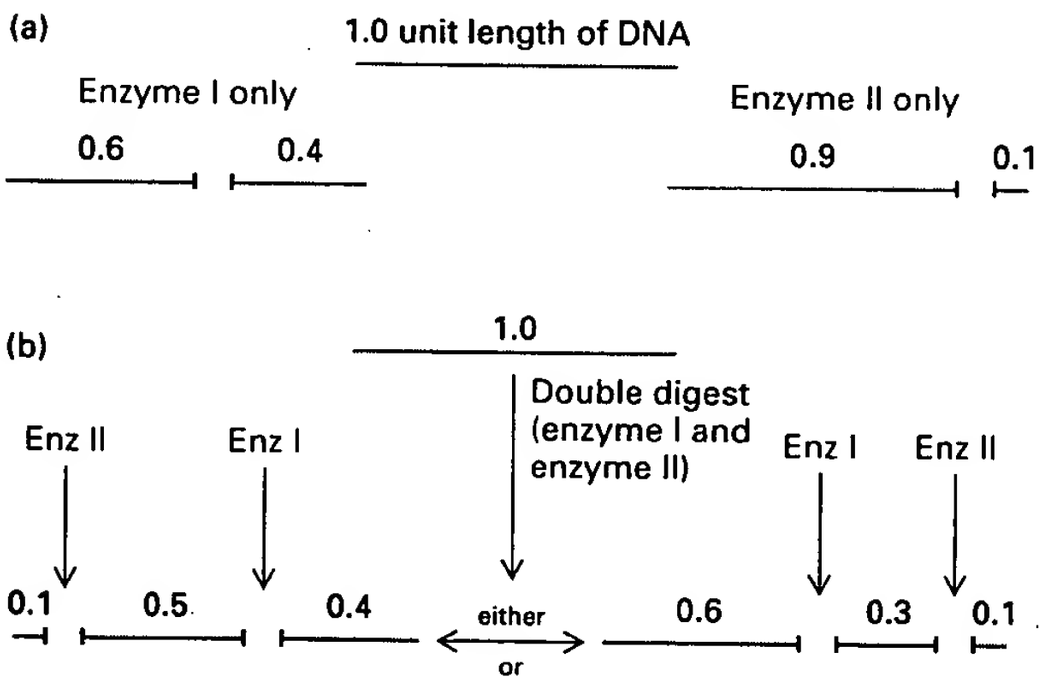
Digestion of DNA by restriction endonucleases, followed by simple electrophoretic separation of the fragments, has revolutionized chromosome mapping. The use of two or more restriction endonucleases on a pure DNA sample can show the order of the restriction sites in a DNA sample (Figure 6-22). Also, many sites can be located by partial digestion of terminally labeled DNA with only one enzyme (Figure 6-23). In these ways, it is possible to produce a map showing the order of the restriction sites in any region of DNA. An important application of restriction endonucleases is their use to cut off one end of a DNA sample that has been end-labeled so that DNA pieces labeled at the other end are available for further study (see Figure 6-23).

Southern DNA Blots The ability to divide DNA into reproducible pieces allows the restriction sites around a particular sequence of interest to be mapped. This possibility is realized in the laboratory by determining which restriction fragments hybridize to a specific labeled *probe* sequence technique called the *Southern blot* (after its originator, Edward Southern). DNA restriction fragments from a sample are separated by gel electrophoresis; their distribution in the gel is preserved as they are denatured and transferred by blotting to a solid substrate with a charged surface (usually a nitrocellulose filter). The filter is then exposed to a specific radioactive nucleic acid sequence (the probe). The blotted DNA fragments that are complementary to the probe hybridize with them, and their location on the filter can be revealed by autoradiography (Figure 6-24). This technique is so sensitive that a DNA sequence that appears only once in the human genome (about 1 part in 10^6) can be detected in as little as 5 μg of DNA (the DNA content of about 10^6 cells).

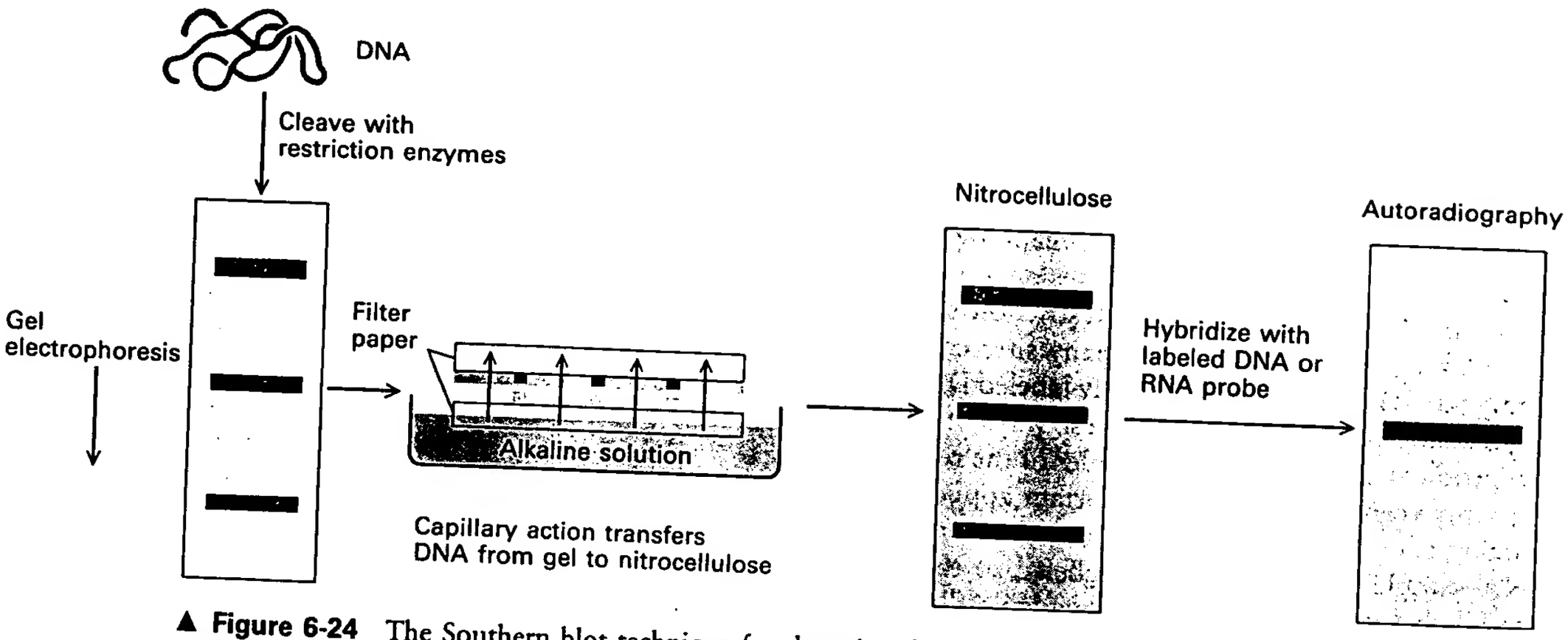
This test is widely used in genetic studies of humans, who do not, as a rule, breed within families. Consequently, the human population shows many genetic differences, or *genetic polymorphisms*. These variations are



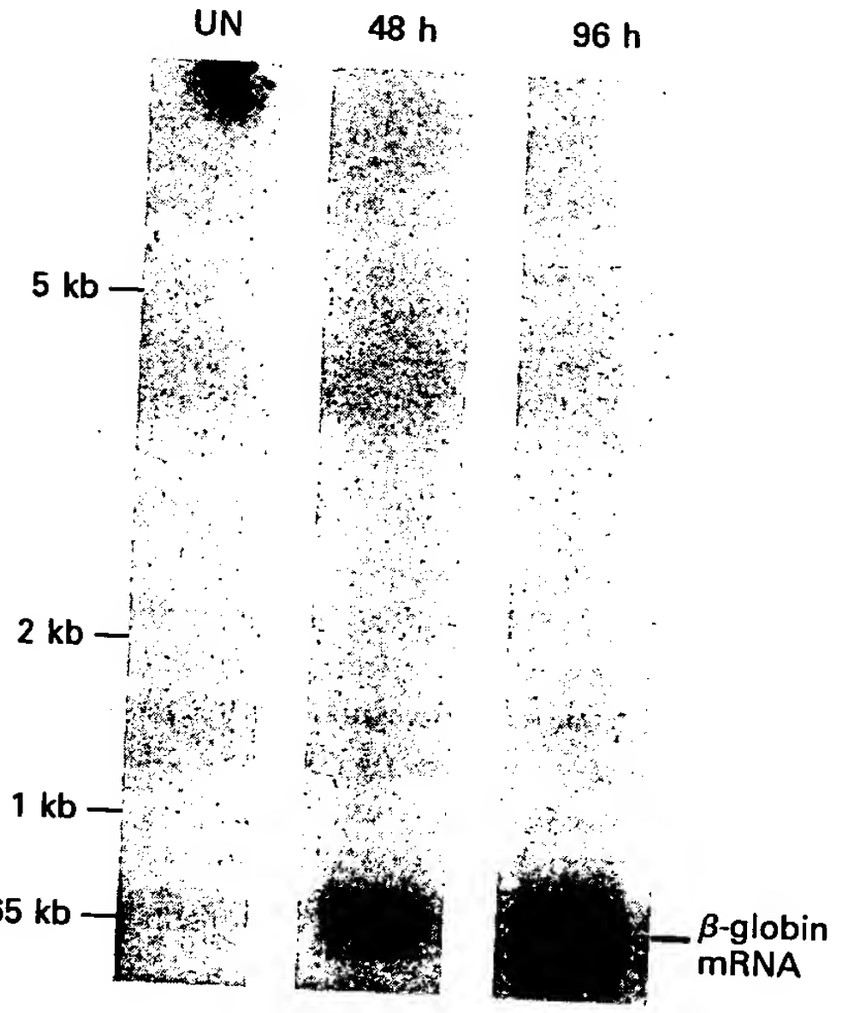
▲ **Figure 6-23** Mapping the multiple recognition sites of a restriction enzyme by partial digestion. DNA is labeled at its termini with ^{32}P, and fragments with *one* labeled terminus can be obtained by cutting off one end with an appropriate enzyme. The mapping procedure is applied to the remaining piece with a second enzyme. Complete digestion would produce only one labeled fragment (here, the 0.2-unit piece), but brief, partial digestion (in which the enzyme cuts each long piece only once, at most) produces a labeled fragment for each restriction site. From the lengths of the labeled pieces, the positions of enzyme II restriction sites can be inferred. [See H. O. Smith and M. Birnstiel, 1976, *Nuc. Acids Res.* 3:2387.]



▲ **Figure 6-22** Mapping the cleavage sites of two restriction enzymes with respect to one another. (a) When a given piece of DNA is exposed separately to two restriction enzymes (I and II), each cuts the DNA once. The lengths of the fragments are determined by gel electrophoresis. (b) Digestion with *both* enzymes is used to determine the relative positions of the cuts along the DNA. The fragment lengths identify the positions of the restriction sites for enzymes I and II with respect to the ends of the DNA and therefore with respect to each other. By continuing this process with different pairs of enzymes, the investigator can construct a detailed map of restriction sites.



▲ **Figure 6-24** The Southern blot technique for detecting the presence of specific DNA sequences. [See E. M. Southern, 1975, *J. Mol. Biol.* 98:508.]



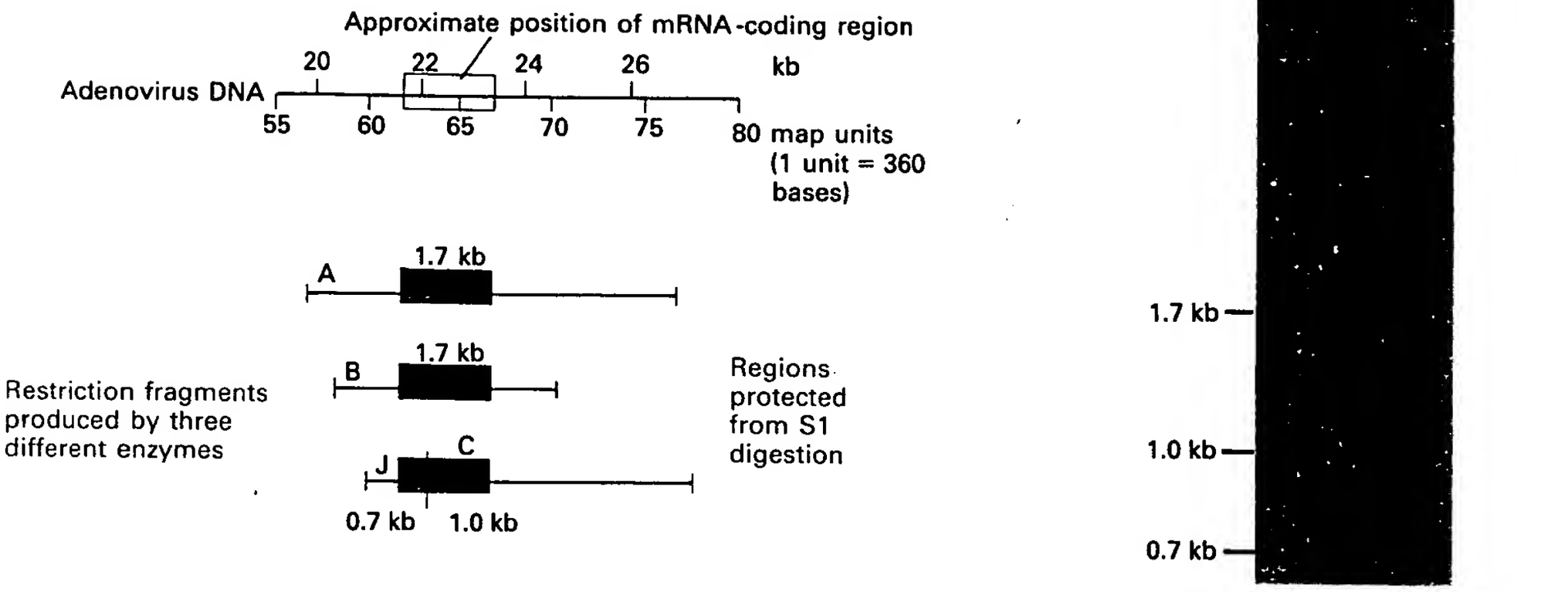
▲ **Figure 6-25** The Northern blot technique for detecting the presence of specific mRNA molecules. Autoradiography shows the position of the complementary mRNA in the gel, and the density of the spot shows the amount of it. The photograph indicates the relative quantities in kilobases (kb) of β-globin mRNA in erythroleukemia cells at three different times: when cells are growing and have not started to make globin (lane UN, for "*un*induced"), and 48 and 96 h after they have been induced to stop growing and begin differentiating. The β-globin mRNA is barely detectable in growing cells but increases by a factor of more than 1000 in 96 h of differentiation. *Photograph courtesy of L. Kole.*

often indicated by the presence or absence of particular restriction sites in the DNA, called *restriction fragment-length polymorphisms*.

Northern (RNA) and Western (Protein) Blots The *Northern blot*, so-named because it is patterned after the Southern blot, is used to detect the presence of specific mRNA molecules. The RNA molecules in a sample are denatured by mixing them with an agent, such as formaldehyde, to prevent hydrogen bonds between base pairs (stems) and ensure that the RNA is in unfolded, linear form. The RNA sample (often the total RNA from cells) is then separated according to size by gel electrophoresis; as in a Southern blot, it is then transferred to a nitrocellulose filter to which the extended denatured RNA will adhere. The filter is exposed to a labeled DNA probe and subjected to autoradiography. The Northern blot indicates the amount as well as the presence and size of a specific mRNA in a sample and the procedure is widely used to compare the amounts of a specific mRNA in cells under different conditions (Figure 6-25).

Another bit of laboratory jargon that has become a widely accepted name for a common technique is the *Western blot*. In this procedure, a one- or two-dimensional electrophoretic separation of proteins is carried out and the protein is then transferred, or blotted, to nitrocellulose. The nitrocellulose sheet can be exposed to radioactive antibodies against a particular protein; autoradiography reveals the presence of that protein.

Band Analysis of S1 Digests An important method for measuring the length of complementary sequences in two nucleic acids employs the endonuclease S1, an en-

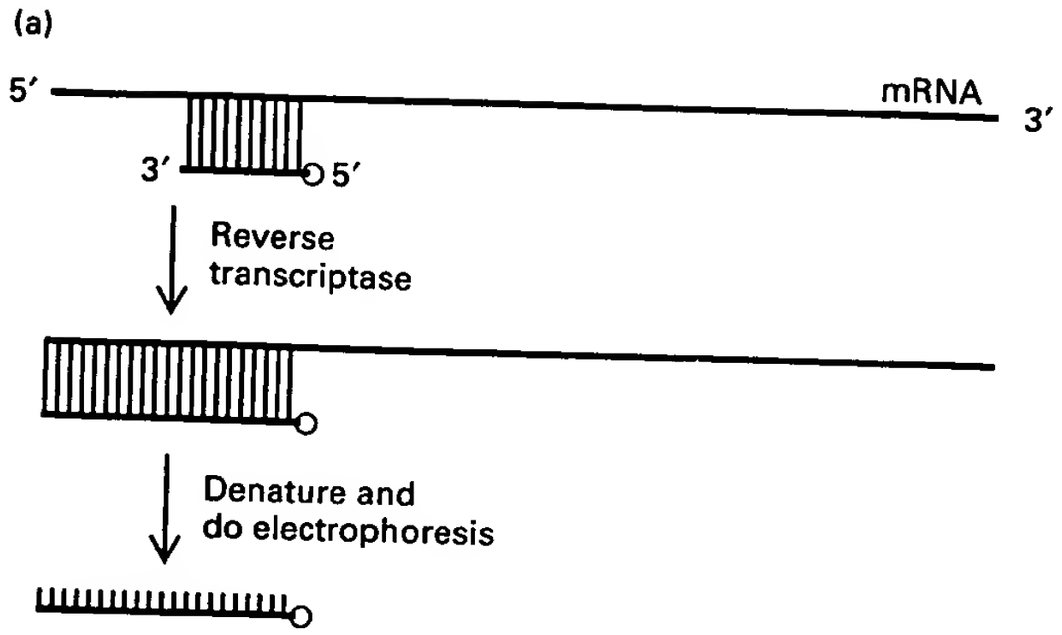


▲ **Figure 6-26** The S1 mapping technique determines the lengths of complementary sequences in two nucleic acid samples. A portion of the map of adenovirus DNA is shown. Earlier hybridization experiments established that an mRNA was complementary to a sequence in the large region spanned here by restriction fragment A. Restriction fragments A, B, C, and J were prepared from ^{32}P-labeled DNA, denatured and hybridized with a large excess of mRNA prepared from virus-infected cells. The mixture was treated with S1 to destroy any unpaired DNA that had not found an mRNA partner; the protected labeled DNA fragments (red) were then separated by gel electrophoresis. The autoradiograph of the gel shows the lengths of the segments protected by (complementary to) the mRNA. Fragments A and B were protected for 1.7 kb; C and J were protected for 1.0 and 0.7 kb. Thus the mRNA includes 1.7 kb positioned as indicated. *From A. J. Berk and P. A. Sharp, 1978,* Cell *14:695. Copyright M.I.T.*

zyme from the mold *Aspergillus oryzae* that destroys unpaired RNA or DNA but not double-stranded molecules (Figure 6-26). Either the RNA or the DNA in a hybrid may be labeled. For example, the total unlabeled mRNA from a cell can be exposed to a labeled DNA probe (usually consisting of one or more restriction fragments) that may include all or part of the region of DNA that is transcribed to produce one particular mRNA. The labeled RNA-DNA hybrid is then digested with S1 to remove unpaired nucleic acid strands, leaving hybrid duplexes intact. After electrophoresis, different hybrids form discrete bands, whose positions can be used to estimate the lengths of the hybrids. This technique is widely used to determine how much of a particular DNA restriction fragment is complementary to an mRNA region.

Finding the Start Site of an mRNA It is often very important to find the point in a DNA sequence at which transcription of a particular mRNA begins. Two methods are used: the endonuclease S1 or the *primer extension technique*. First, a general region of a DNA molecule that includes the start site must have been located (DNA sequencing is described in the next section). Appropriate restriction sites can be chosen in this sequence to prepare a piece of end-labeled DNA approximately 100 nucleotides long that will hybridize with the 5′ portion of the mRNA. Figure 6-27 shows how such a piece of labeled DNA, trimmed with S1 endonuclease, can be used to locate the exact start site. The same logic applies with primer extension. A DNA oligonucleotide about 20 bases long is chosen to find a specific complementary site on the mRNA. This primer can be extended enzymatically to the beginning of the mRNA; the length of this extension product can then be determined accurately by gel electrophoresis.

Endonucleases Compared to Exonucleases Thus far, we have been discussing endonucleases, enzymes that cut DNA or RNA (or both) *within* a chain. Restriction endonucleases have a restricted cutting specificity; S1 and the widely used pancreatic RNase and DNase digest almost all internucleotide bonds equally. The S1 endonuclease is single-stranded; the pancreatic enzymes can be single- or double-stranded.

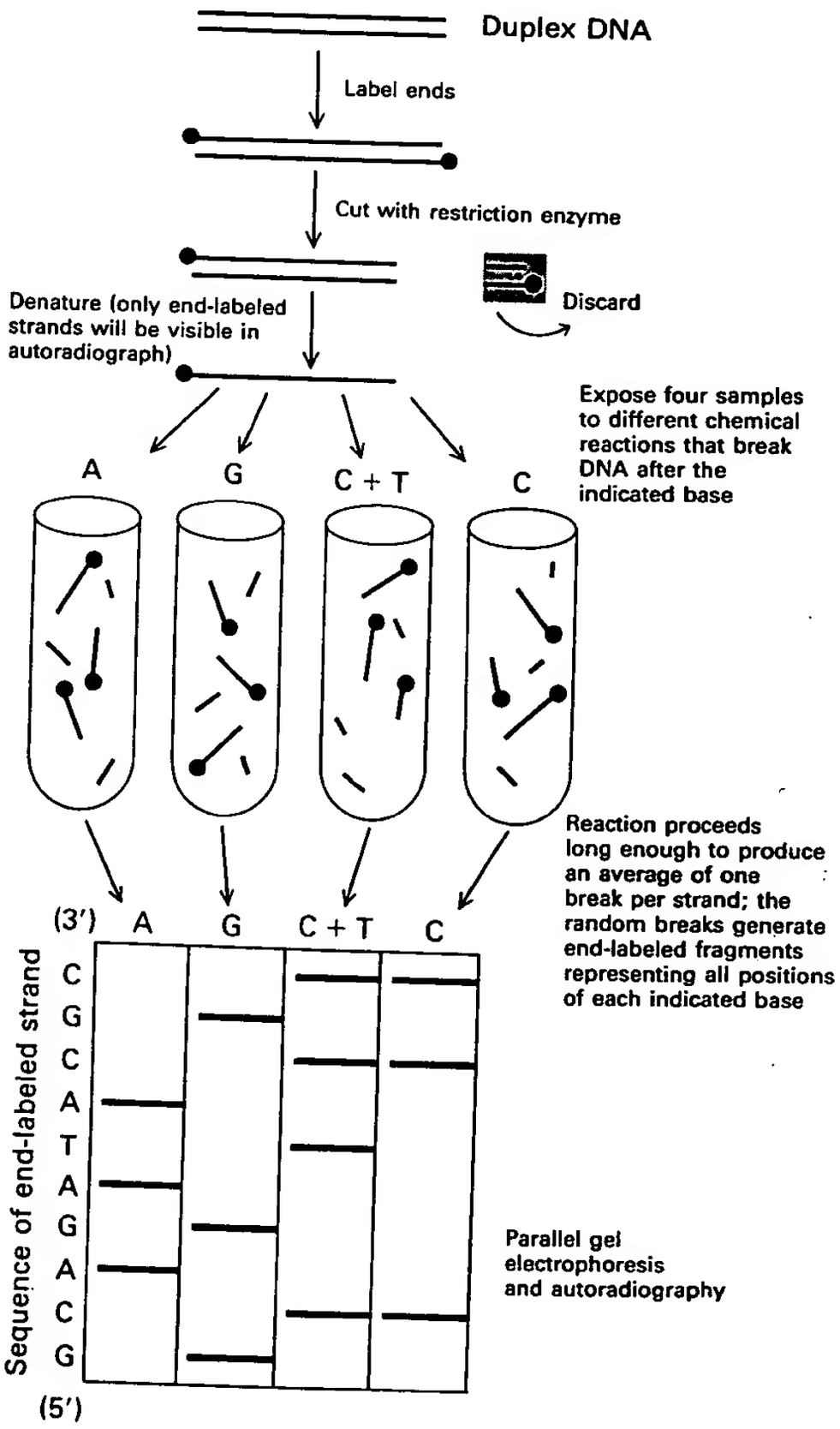


▲ **Figure 6-27** The site in DNA that encodes the first nucleotide in an mRNA molecule can be found by using primer extension or S1 endonuclease (see Figure 6-26). (a) In the primer extension technique, a short (approximately 10 nucleotides) oligodeoxynucleotide (blue) is prepared and end-labeled. After it is hybridized to the mRNA (red), it is lengthened by the enzyme reverse transcriptase until it reaches the first nucleotide of the mRNA. (b) The use of S1 to map start sites begins with the preparation of a uniquely end-labeled, short (approximately 100 nucleotides), single DNA strand, that encodes the general region of the mRNA start site and whose total sequence is known. This is hybridized with the mRNA, and unpaired nucleic acid is then digested with S1 endonuclease. Denaturation leaves a labeled DNA piece whose length accurately marks the distance of the starting nucleotide of the mRNA from the nucleotide that hybridized with the labeled DNA end.

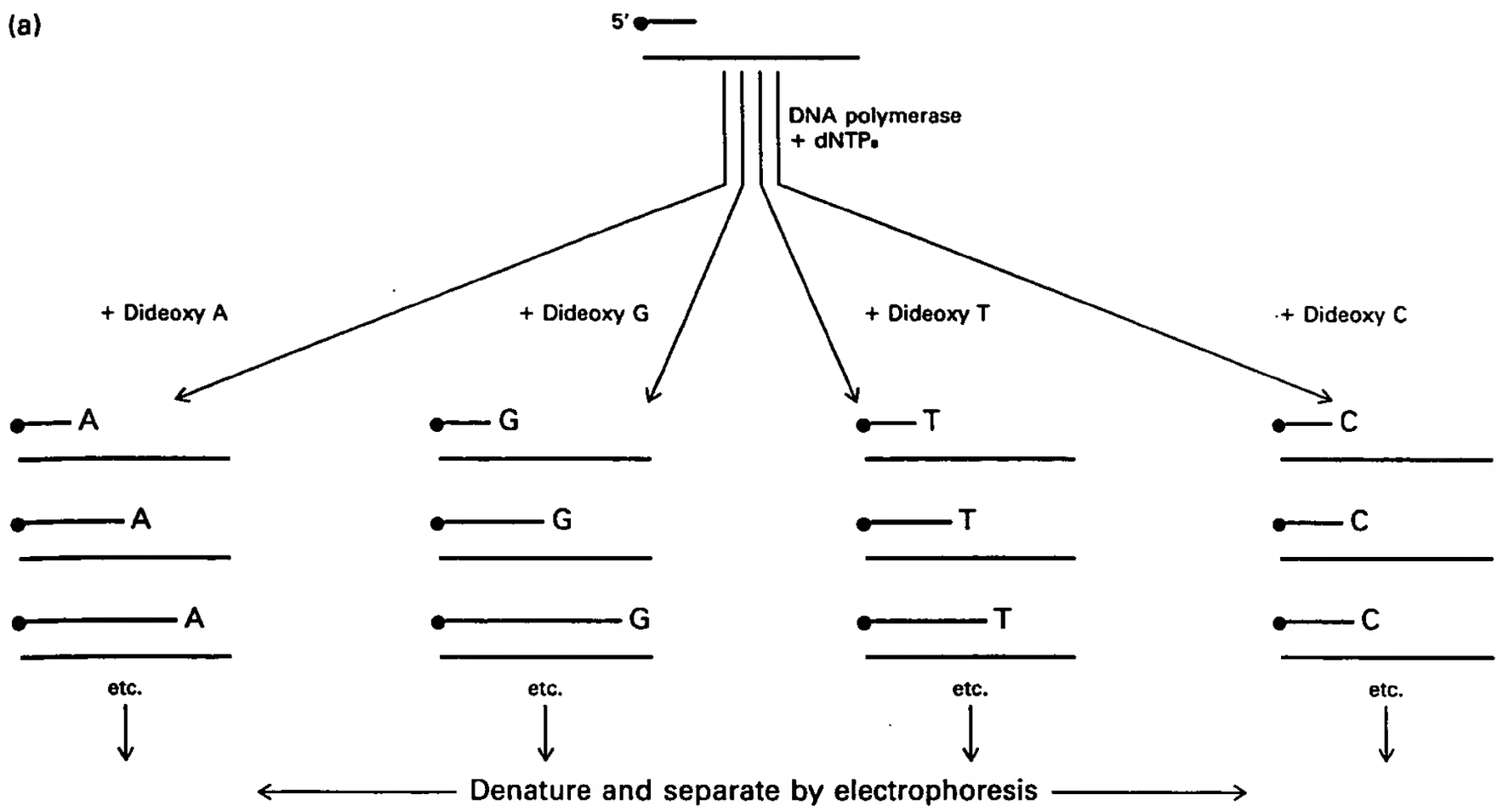
Other enzymes called *exonucleases* remove nucleotides one at a time from the *ends* of entire RNA or DNA strands. Some act only on single strands; others remove nucleotides from either the 5′ or the 3′ ends (but not both) of duplex DNA. There are too many of these enzymes to attempt a comprehensive description of them here. We shall describe exonucleases as necessary in this and other chapters.

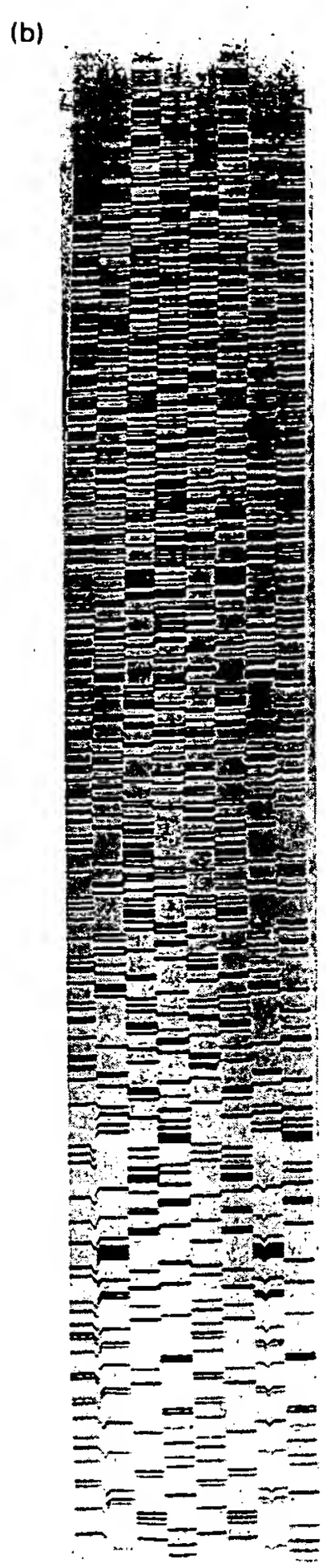
The Sequence of Nucleotides in Long Stretches of DNA Can Be Rapidly Determined

The discovery of restriction endonucleases was an important step that led to general methods for determining the exact nucleotide sequences in long stretches of DNA. An



▶ **Figure 6-28** DNA sequencing by the Maxam-Gilbert method. A 5′-end–labeled DNA fragment is prepared for sequencing. Four identical samples of this fragment are subjected to four different chemical reactions. Each breaks the fragment only (or mainly) at the A, G, C + T, or C residues, respectively. The reactions are controlled, so that each labeled chain is likely to be broken only once. The resulting *labeled* subfragments created by all four reactions have the label at one end and the chemical cleavage point at the other. Gel electrophoresis and autoradiography of each separate mixture yield one radioactive band for each nucleotide in the original fragment. Bands appearing in the A and G lanes can be read directly. Bands in the C + T and C lanes are read as Cs; those in the C + T lane alone, as Ts. [See A. Maxam and W. Gilbert, 1977, *Proc. Nat'l Acad. Sci. USA* 74:560.]





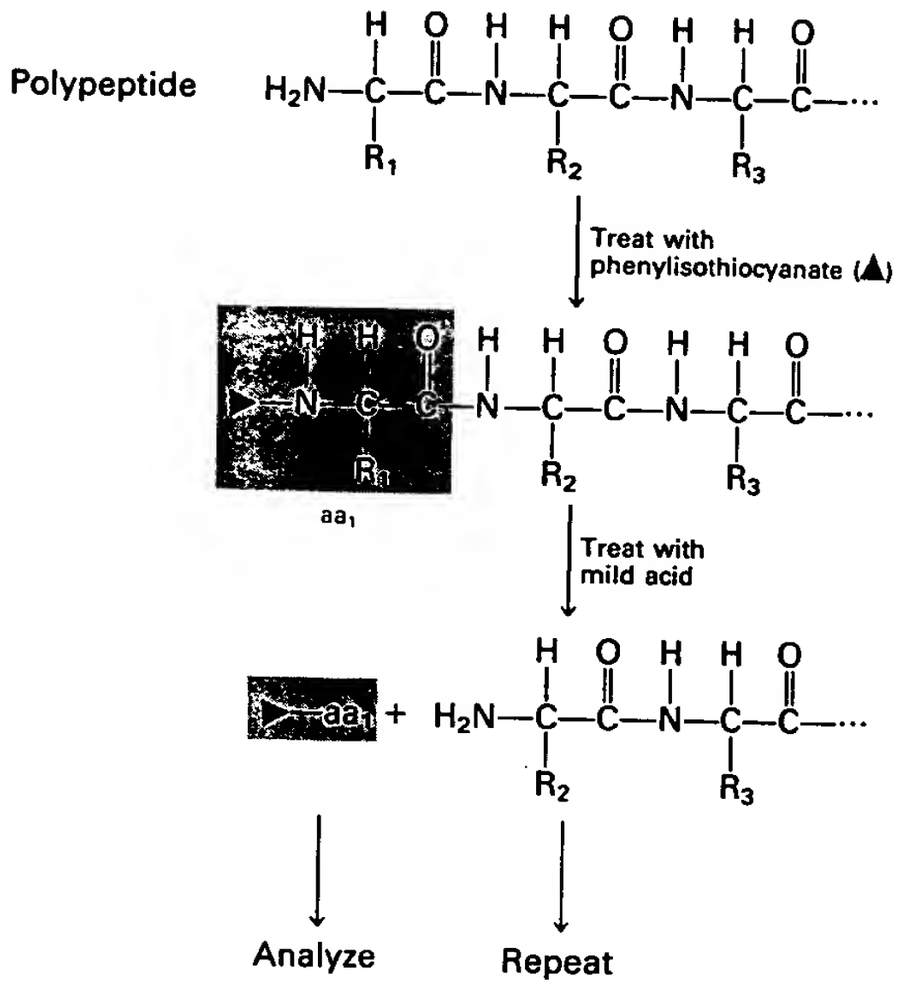
▲ **Figure 6-29** DNA sequencing by the Sanger (dideoxy) method. (a) A single strand of DNA to be sequenced is hybridized to a 5′-end–labeled deoxynucleotide primer; four separate reaction mixtures are prepared in which the primer is elongated by a DNA polymerase. Each mixture contains the four normal deoxynucleoside triphosphates plus one of the four dideoxynucleoside triphosphates in a ratio such that about 1 in every 200 residues is a dideoxynucleotide. Since a dideoxynucleotide has no 3′ hydroxyl, no further chain elongation is possible when such a residue is added to the chain. Thus, each reaction mixture will produce prematurely terminated chains ending at every occurrence of the dideoxynucleotide. Each mixture is then separated on a sequencing gel as in Figure 6-27. (b) An actual radioautogram in which over 300 bases can be read. (The enzyme used in the figure is Sequenase™, a commercial preparation of a bacteriophage-encoded polymerase.) Each reaction was carried out in duplicate, which aids in checking the sequence. *Courtesy of United States Biochemical Corporation.*

earlier and very important advance was the recognition that careful gel electrophoretic procedures can separate every single DNA fragment in a series up to 500 bases long. Two highly successful procedures for DNA sequencing are in wide use. Both depend on the separation of a set of fragments that differ in length by only one nucleotide (Figure 6-28). The first, invented by Allan Maxam and Walter Gilbert, chemically cleaves an end-labeled DNA sample; the second, developed by Fred Sanger and his colleagues, uses enzymatic synthesis to extend a short sequence of end-labeled DNA (Figure 6-29).

Modern DNA sequencing is fairly simple and accurate over long regions; already, the total genomes of many viruses and almost all of the *E. coli* genome have been sequenced. Automation of the techniques for sequencing large pieces of DNA (see Figure 6-8) should permit sequencing of the entire human genome in 10–15 years, if not before.

Proteins Can Be Sequenced Automatically

From a DNA sequence and the genetic code, it is possible to deduce the sequences of the encoded protein. And with the aid of computers to locate "open reading frames," i.e., codon stretches without protein termination signals, investigators often do just that. Nevertheless, the ability to sequence protein chains directly remains a crucially important and necessary tool of molecular biology. To cite one application, the genome of a higher animal may contain a number of genes that are similar but not identical in sequence; only by knowing the protein sequence of a product of such related regions can the observer know which DNA sequence is responsible for encoding a particular protein. Even more importantly, perhaps, proteins of interest are most often isolated *before* their genes, so obtaining at least a partial amino acid sequence is a critical first step in studying many proteins. The most popular direct protein-sequencing technique in use today is the



▲ **Figure 6-30** Sequencing a protein by the Edman degradation procedure. The peptide is treated with phenylisothiocyanate, which combines with the amino-terminal residue in the peptide chain, rendering the first peptide bond in the chain labile to treatment with mild acid. The same pair of reactions is carried out repeatedly to remove the amino acids one at a time. After each step, the removed amino acid is chemically identified. In this way, the entire amino acid sequence of a short peptide can be determined.

Table 6-5 Terms used in recombinant DNA research

Genomic DNA	All DNA sequences of an organism
cDNA (complementary DNA)	DNA copied from an mRNA molecule
Plasmid	A small, circular, extrachromosomal DNA molecule capable of reproducing independently in a host cell
Vector	A plasmid or a viral DNA molecule into which either a cDNA sequence or a genomic DNA sequence is inserted
Host cell	A cell (usually a bacterium) in which a vector can be propagated
Genomic clone	A selected host cell with a vector containing a fragment of genomic DNA from a different organism
cDNA clone	A selected host cell with a vector containing a cDNA molecule from another organism
Library	A complete set of genomic clones from an organism or of cDNA clones from one cell type

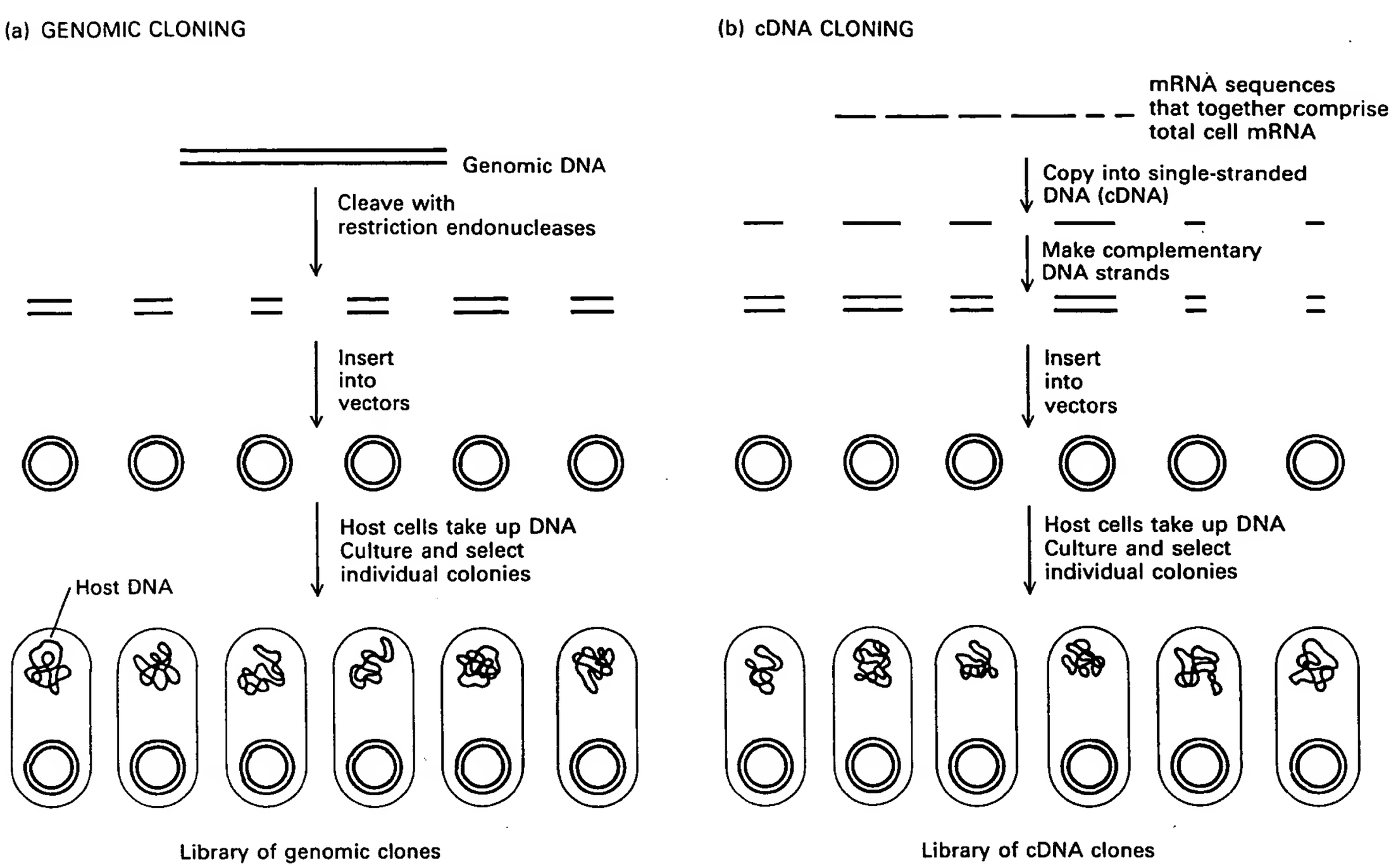
Edman degradation procedure, in which amino acid residues are cleaved from a protein one by one; after each cleavage, the released amino acid is identified (Figure 6-30). Machines called sequenators can perform this reaction on tiny amounts of a pure protein; obtaining an accurate sequence of 50 amino acids is not exceptional.

Recombinant DNA: Selection and Production of Specific DNA

The essence of cell chemistry is to purify sufficient quantities of a particular substance to permit its chemical behavior to be analyzed. Segments of pure samples of identical, relatively short DNA molecules from viruses or plasmids can be isolated directly and subdivided into smaller pieces with the use of restriction endonucleases. But the human genome, for example, contains about 3×10^9 bp, so that cutting roughly at every 3000th base pair would produce a million fragments that could not be separated from each other. This obstacle to obtaining pure DNA samples from large genomes has been overcome by recombinant DNA technology.

Two widely used types of recombinant DNA preparations—genomic clones and cDNA clones—are made. A *genomic clone* contains a fragment of genomic DNA; a *cDNA clone* contains a molecule of *complementary DNA* copied from mRNA by enzymes (Table 6-5). In both, the DNA of interest is linked to a *vector*—most often a bacteriophage or a plasmid that can reproduce independently within a bacterial host. (The most widely used *host-vector systems* are E. coli as host with either a plasmid or bacteriophage λ as the vector.) Recently yeast artificial chromosomes (YACs) have been prepared that can be used as vectors in yeast cells for very large genomic fragments. A *library* consisting of a full set of genomic or cDNA clones can be prepared from the total DNA of an organism or cell type or from the set of cDNA molecules copied from all mRNAs in a cell (Figure 6-31).

The preparation and selection of cDNA and genomic clones are illustrated in the following section by a description of how recombinant DNA containing mouse globin sequences can be obtained.



▲ **Figure 6-31** A comparison of genomic cloning (a) with cDNA cloning (b). In genomic cloning, the genomic DNA must be cleaved with restriction endonucleases before it can be inserted into vectors; in cDNA cloning, the mRNAs must first be copied into double-stranded DNA molecules.

cDNA Clones Are Whole or Partial Copies of mRNA

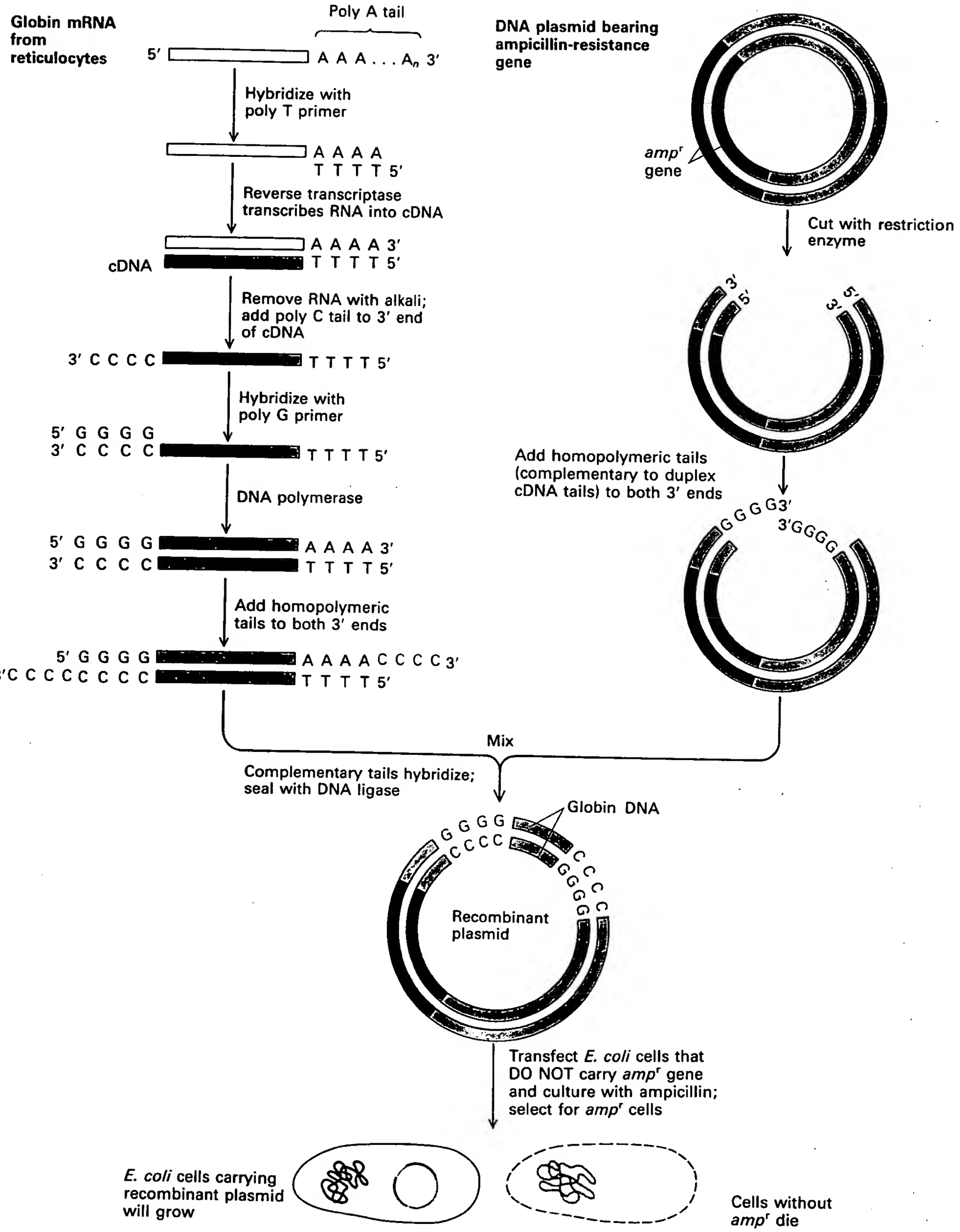
To prepare cDNA clones with globin-encoding sequences (Figure 6-32), the starting material is reticulocytes, erythrocyte (red blood cell) precursors. Over 90 percent of the proteins synthesized by these cells are α- and β-globins and therefore they are rich sources of globin mRNA.

The enzyme *reverse transcriptase* (found in retroviruses; see Figure 5-39) is used to make cDNA clones of the reticulocyte mRNAs. Like the DNA polymerases in cells, this enzyme can build a complementary nucleic acid strand on a template, but only by adding nucleotides to a primer. Thus, before the reverse transcriptase can do its work, a short primer strand must be hybridized to the nucleotides near the 3′ ends of the mRNAs. Fortunately, a single oligonucleotide primer—a string of thymidylate residues (poly T)—serves for most eukaryotic mRNAs, which end in a string of 50–250 adenylate residues (poly A).

After the cDNA copy of the mRNA has been made, the mRNA is removed by an alkali treatment that destroys RNA but does not affect DNA and a duplex DNA is made from the cDNA strand. In one technique, the 3′ end of each cDNA strand is elongated by adding several residues of a single nucleotide (say, poly C) through the action of a *terminal transferase,* an enzyme that adds bases at free 3′ ends. A poly G primer is hybridized with the terminal poly C and this G primer is then elongated by a DNA polymerase. What results is a complete double-stranded DNA copy of the original mRNA.

The next step is to insert the now double-stranded DNA into a plasmid. Plasmids, which occur naturally in almost all bacteria, were originally detected by their ability to transfer genes between bacteria (Chapter 5). It has been shown that a specific region of the plasmid circle, the *replication origin,* must be present to assure replication of the plasmid in a host bacterium.

The plasmid DNA is cleaved once with a restriction enzyme at a point that leaves the replication origin undisturbed. The double-stranded copy of the mRNA-globin is then inserted at the cut site and the circle is rejoined. The



▲ **Figure 6-32** Preparation of a cDNA clone with globin encoding sequences.

first technique, still widely used, for carrying out this insertion is called *homopolymeric tailing*. A homopolymer (say, poly C) is added to the two 3′ ends of the double-stranded cDNA-globin, and a complementary homopolymer (poly G) is added to the 3′ ends of the cut plasmid. When the "tailed" plasmid and DNA-globin are mixed, their complementary single-stranded tails spontaneously hybridize; the resulting circular recombinant molecule can be resealed with the enzyme DNA ligase (Chapters 3 and 12). Specially treated *E. coli* cells take up the plasmid, and the recombinant molecule multiplies along with the cells.

If the chosen plasmid contains a gene for resistance to an antibiotic, the cells that take up the plasmid will grow and multiply in the presence of the antibiotic but the other cells will not. If, at the outset, the number of plasmids allowed to infect the *E. coli* cells is one-tenth or less of the total number of *E. coli* cells, it is very unlikely that more than one plasmid will end up in a recipient cell. As a rule, then, the recombinant DNA in all cells of a colony grown from a single cell will have descended from a single recombinant DNA molecule. In the case described here, 90 percent of the recombinant plasmids would encode α- or β-globin. Because mRNA molecules are often not completely copied, partial sequences also may be cloned. To verify exactly what the plasmid vector contains, the recombinant molecule can be sequenced.

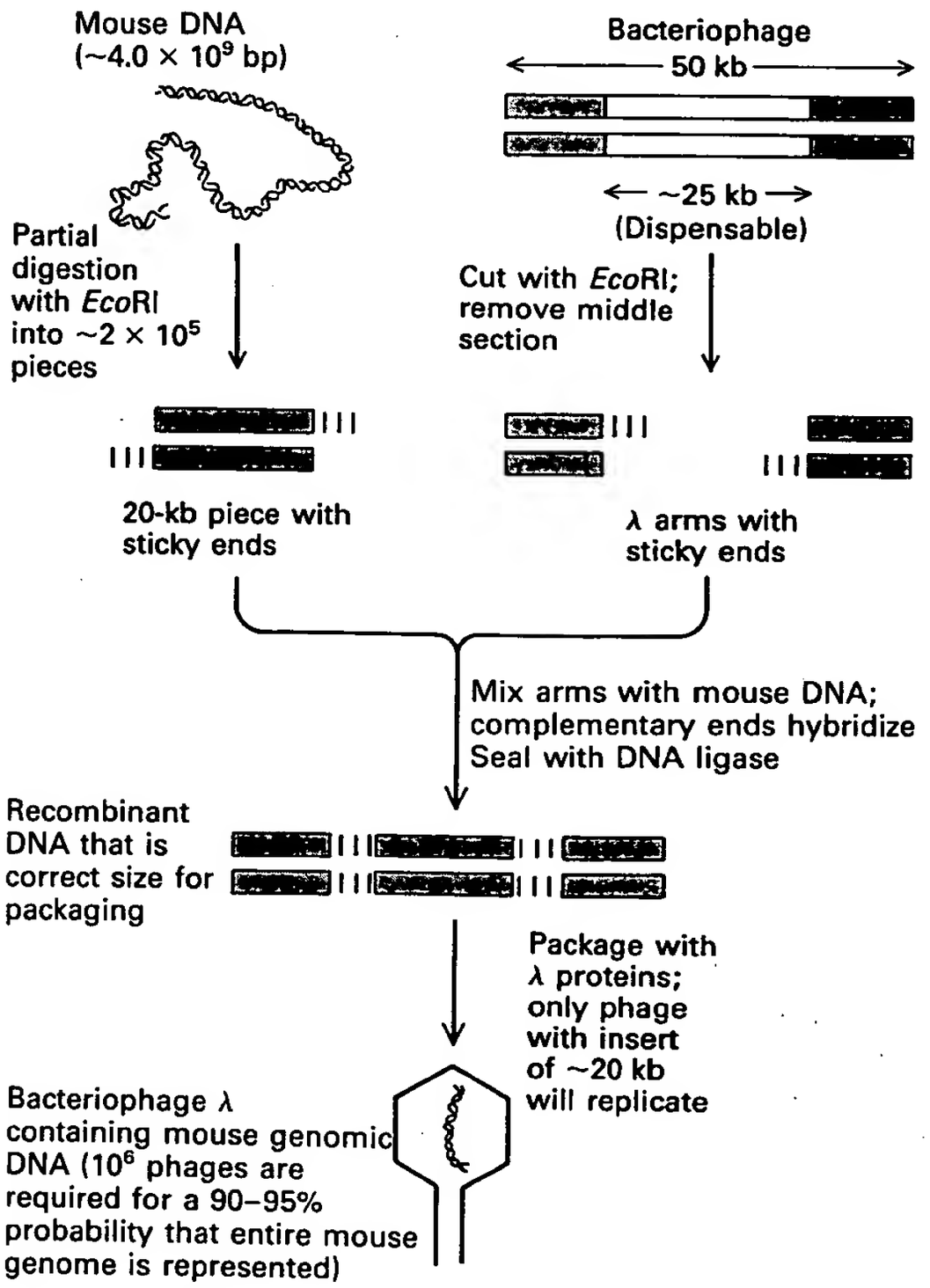
Complementary DNA clones can be prepared from the unpurified mRNA from any cell type, but this produces a random mixture of individual recombinant clones that must be screened to isolate specific ones (see Figure 6-34). It is also possible to use an antibody that reacts with a protein to detect whether an *E. coli* colony (or a plaque if the vector is a bacteriophage) contain the protein encoded by the cloned cDNA.

Genomic Clones Are Copies of DNA from Chromosomes

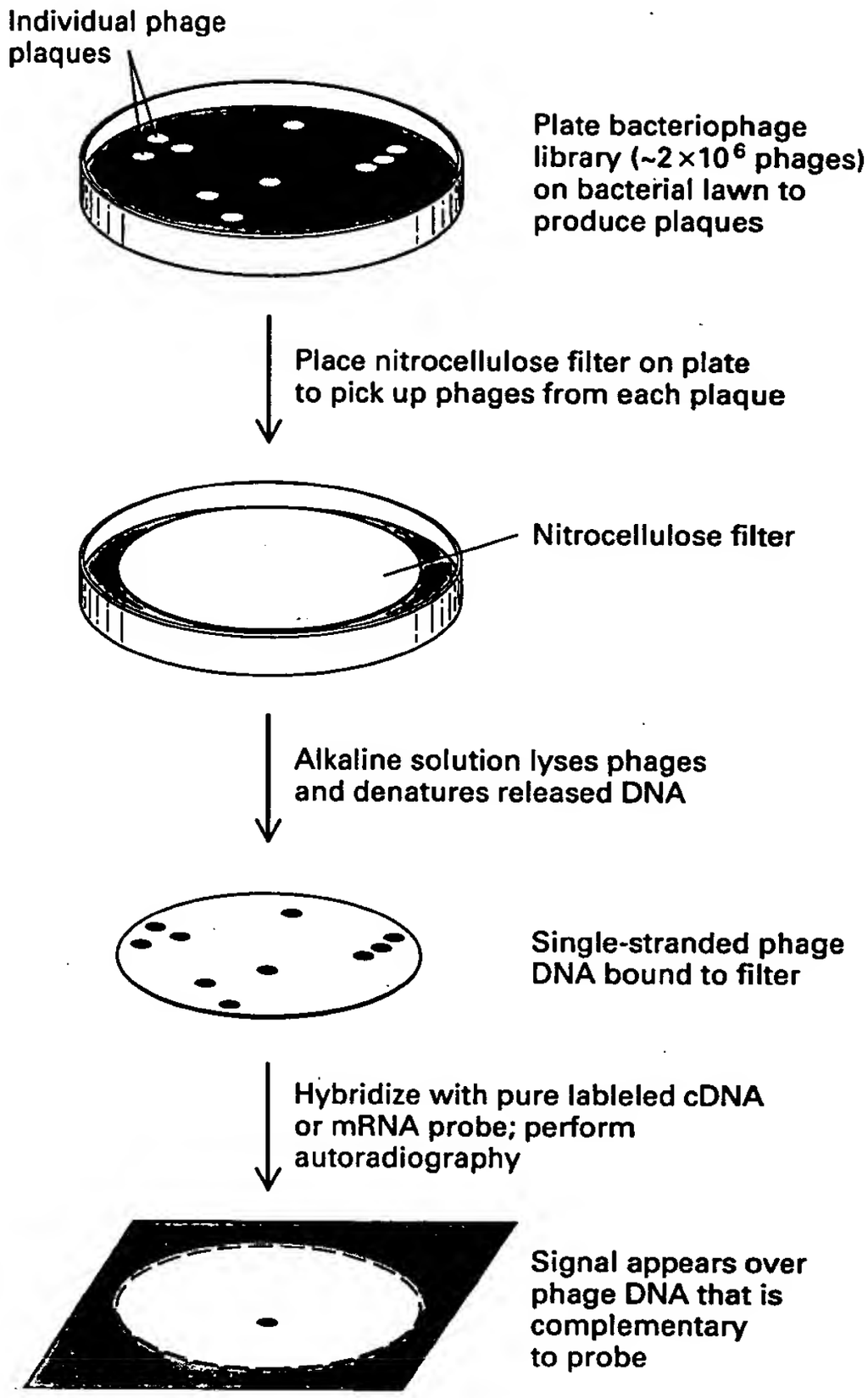
The most common procedure for preparing and selecting specific clones from genomic DNA—for example, the total collection of DNA in mouse chromosomes—makes use of λ bacteriophage. The DNA of the phage is about 50 kb long, but a center section about 25 kb long can be removed and replaced with foreign DNA without impairing the ability of the phage to infect and reproduce in most *E. coli* cells. A genomic library is a collection of recombinant molecules, maintained either in phage particles or in plasmids growing in bacteria, that includes *all* DNA sequences of a given species. Once it is prepared, the library can be screened for the phage or plasmid that contains the DNA sequence of interest.

The size of a library depends on the amount of DNA in the organism's haploid genome. For example, the human and mouse genomes are between 3 and 4×10^9 bp long. If one of these genomes were divided into fragments about 20 kb long for insertion into bacteriophage λ, then 2×10^5 different recombinant bacteriophage λ particles would be required to constitute a complete library. Because the pieces of DNA are incorporated into phages by chance, about 10^6 recombinant phages are necessary to assure that each DNA piece has a 90–95 percent chance of being included.

The first step in preparing a genomic library is to extract all the organism's DNA from some cell types (Figure 6-33). Sperm cells or early embryos are often used. The DNA is then broken into fragments by a restriction endonuclease, such as *Eco*RI, which cleaves the DNA in a way that produces short, single-stranded, "sticky" ends (AATT) on every fragment (see Figure 6-20). Digestion is stopped when the average size of a fragment is approxi-



▲ **Figure 6-33** The construction of a genomic library of mouse DNA in bacteriophage λ. The total DNA from mouse cells (both sperm cells and embryonic tissue cells presumably have a complete set of sequences) is often used. A single region of the mouse genome, such as the one that encodes β-globin, would occur approximately once in 10^5 particles.



► **Figure 6-34** Selection of a specific genomic clone from a bacteriophage λ library. Although about 2×10^5 phages *could* contain all mouse sequences, 2×10^6 phages are plated to ensure that a phage with the desired sequence is included. This requires an area of 1000–2000 cm^2 to accommodate all the phage plaques. (In the initial plating, the plaques are not allowed to develop to a visible size. The plating can be repeated with fewer phages to obtain pure isolates.) The position of the spot on the autoradiograph identifies the desired plaque on the plate. Phage particles from that plaque can then be selected.

mately 20 kb. The bacteriophage λ DNA also can be cut at two restriction sites by *Eco*RI to yield a center section approximately 25 kb long plus two shorter flanking ends, or *arms*. The center section of the phage DNA can then be separated from the two arms.

The λ arms and the collection of genomic DNA fragments are mixed in about equal amounts (approximately 10^6 DNA fragments and a similar number of pairs of λ arms). Because the sticky ends are complementary, molecules approximately the same length as normal phage DNA will form, but they will include a piece of mouse DNA about 20 kb in length. DNA ligase, the enzyme that normally joins DNA breaks, is used to seal the recombinant DNA molecules, which are then coated with bacteriophage proteins prepared from infected *E. coli* cells. Only DNA molecules of the correct size will be effectively coated, or *packaged,* and give rise to fully infectious λ-bacteriophages, which can be grown on a lawn of *E. coli.* Bacteriophages containing DNA sequences that code for any specific sequence (for example, globin) can be detected by hybridization of cDNA-globin sequences (prepared as described in Figure 6-32) with DNA obtained from each plaque (Figure 6-34).

Vectors for Recombining DNA Exist in Many Cell Types

Any gene that can be subjected to a hybridization assay can be purified. Once a bacteriophage or other bacterial vector containing the desired gene is prepared, an unlimited amount of the pure gene can be obtained by growing the vector and extracting the DNA. Vectors can also carry recombinant DNA molecules in yeast, higher plant cells, and human cells. The vectors most frequently used in mammalian cells are the small DNA viruses SV40 and polyoma or the slightly larger papilloma viruses that can grow as plasmids. Retroviruses are reminiscent of transducing bacteriophages in that they enter a cell, their RNA is copied into DNA, and then inserted into the host chromosomes. Defective retrovirus vectors, which can sponsor DNA copying and insertion but cannot reproduce themselves, promise to be a successful means of gene therapy for individuals with single genetic defects that may be treated in somatic (body) tissue. In plant cells, the most common vector is the Ti plasmid, whose host is *Agrobacterium tumefaciens*; this bacterium fuses with and transfers the recombinant DNA to the plant cell.

For the cell biologist, the availability [through the use of recombinant DNA technology] of unlimited amounts of a pure gene offers rich opportunities for chemical and biological study. Access to vectors in yeast and in cultured mammalian cells affords the additional possibility of testing the biological functions of particular eukaryotic DNA sequences in a variety of eukaryotic cells.

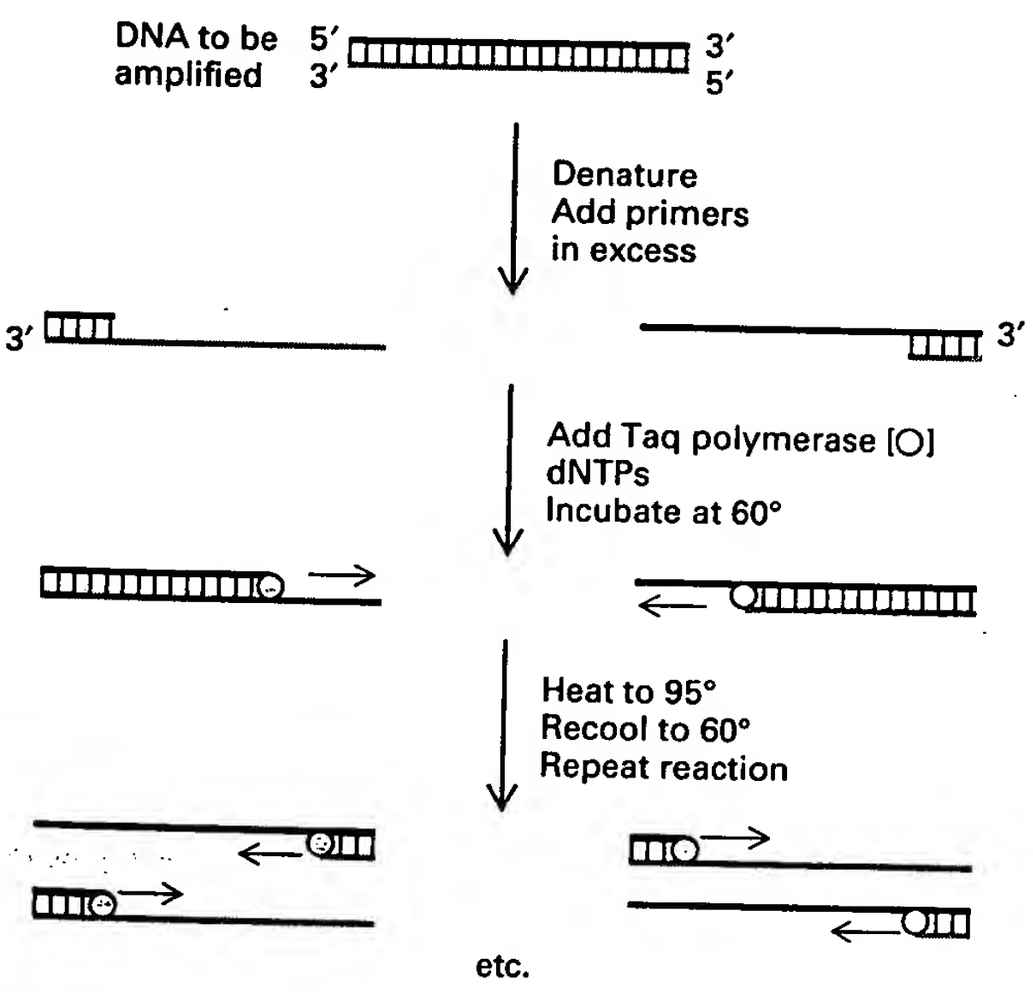
Industrial microbiologists employ recombinant DNA techniques to engineer bacteria and other easily cultured organisms to make proteins for use in medicine, agriculture, and research. A number of viral proteins important

in immunizations (for example, against the foot-and-mouth virus in cattle or the hepatitis virus in humans) have already been synthesized in *E. coli,* as have several hormones and enzymes (among them, insulin, growth hormone, and tissue plasminogen activator [TPA], which is used to combat heart attacks). The vectors that direct such programmed protein synthesis, called *expression vectors,* allow the experimenter to take advantage of bacterial genetic tricks that increase mRNA synthesis to produce large quantities of a desired protein.

The Polymerase Chain Reaction Amplifies Specific DNA Sequences in a Mixture

A new procedure called the *polymerase chain reaction* (PCR) can selectively and repeatedly replicate selected segments from a complex DNA mixture (Figure 6-35). This way of amplifying rare sequences from a mixture has vastly increased the sensitivity of genetic tests.

In a typical application of PCR, DNA from a small sample of blood is cut into segments with a restriction endonuclease and denatured into single strands. Oligonucleotide probes complementary to the 3′ ends of the DNA segment to be amplified are prepared. The probe is added in great excess to the denatured DNA at a temperature between 50° and 60°. The total genomic DNA sample, which is at a low concentration, remains denatured but the specific oligonucleotide probe hybridizes with its correct site on the DNA. The hybridized probe will then serve as a primer for DNA chain synthesis, which begins upon addition of a supply of deoxynucleotides and a temperature-resistant DNA polymerase obtained from *Thermus aquaticus* (a bacterium that lives in hot springs). This enzyme (called the *Taq polymerase*) can extend the primers at high temperatures (up to 72°). When synthesis is complete, the whole mixture is heated further (to 95°) to melt the newly formed DNA duplexes. When the temperature is lowered again, another round of synthesis can take place because excess primer is still present. This cycle of synthesizing and remelting can be repeated to amplify the sequence of interest. At each round, the number of copies of the sequences between the primer sites is doubled and therefore the desired sequence increases exponentially.



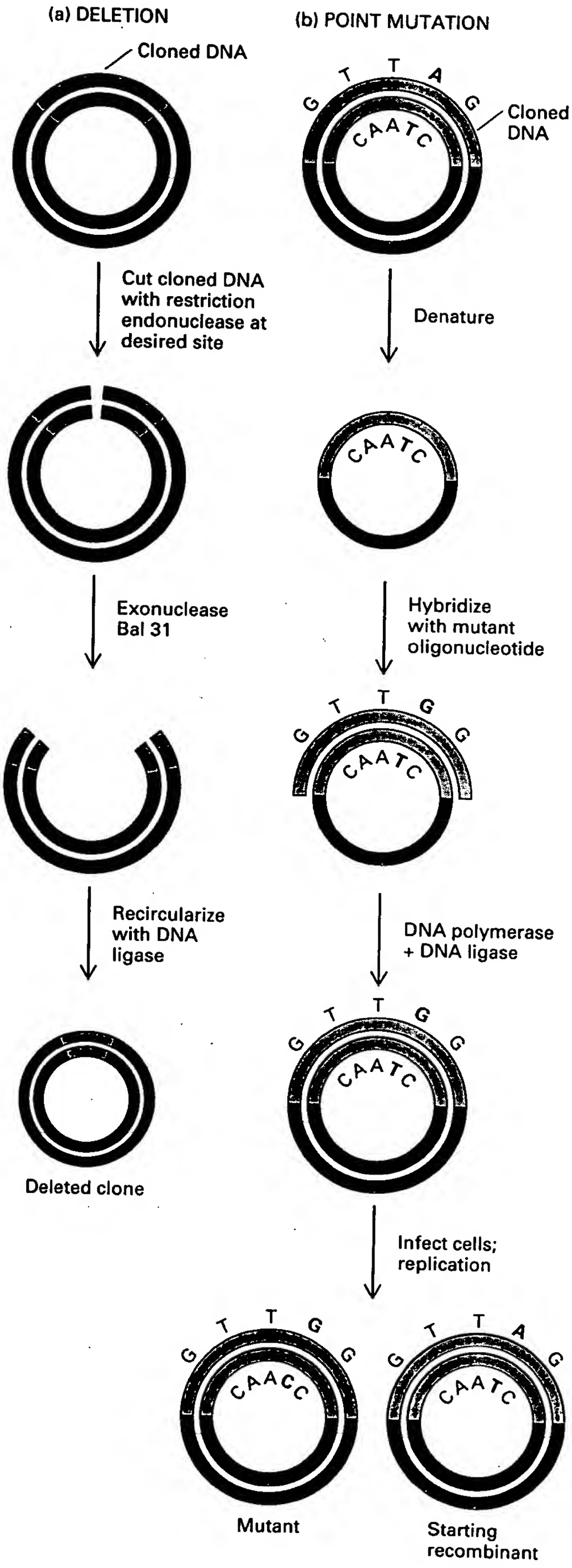
▲ **Figure 6-35** The polymerase chain reaction (PCR). *Taq polymerase,* a heat-resistant DNA polymerase from *Thermus aquaticus,* is used to extend primers between two fixed points on a DNA molecule. All the components for chain elongation (primers, deoxynucleotides, and polymerase) are heat-stable. Thus multiple heating and cooling cycles result in alternating DNA melting and synthesis. DNA between the recognition sites of the two oligonucleotide primers accumulates exponentially. Overnight, it may be amplified as much as a millionfold.

The polymerase chain reaction allows specific DNA regions from a tiny sample to be examined quickly. PCR is already in use as a diagnostic procedure in human genetics. In basic research, PCR allows recovery of entire sequences between any two ends whose sequences are known.

Controlled Deletions and Base-Specific Mutagenesis of DNA

The availability of pure DNA in unlimited amounts has permitted a variety of chemical and enzymatic techniques for altering DNA to be developed. The practice of genetics no longer depends on isolating naturally occurring mutant organisms; DNA can be changed in the test tube and reinserted into cells. Thus deletions and mutations can be introduced into genes. Determining the effects on protein structure and changing DNA sequences that may function as genetic regulatory or control elements are two of the most important uses of these techniques.

Two techniques for introducing mutations—the deletion of a short DNA sequence and the alteration of a single base—are illustrated in Figure 6-36. The function of the mutant DNA—whether it is a *deletion mutant* (Figure 6-36a) or a *point mutant* (Figure 6-36b)—can be tested by reintroducing it into a cell by injection or transformation (Chapter 5). The power of this approach is that without knowing the role of a particular sequence beforehand, the experimenter can determine its function by altering its structure and reintroducing it into the organism. Charles Weissman has termed these practices "reverse genetics."



Synthetic Peptide and Nucleotide Sequences: Their Use in Isolating and Identifying Genes

As more and more primary sequences of proteins and nucleic acids become known, the special importance of certain short sequences—regulatory signals in nucleic acids and functional subsections, or "domains," in proteins—become more apparent. These sequences can be chemically synthesized. With such fragments, the function of a part of a protein, rather than the whole protein, can be tested or altered oligonucleotides can be inserted into normal cloned DNA sequences to study the effects of specific mutations (see Figure 6-36).

Another extremely valuable aspect of synthetic oligonucleotides and peptides is that they make it possible to isolate whole genes and pure proteins, respectively. Because the genetic code is universal, a nucleic acid sequence can be used to predict the exact protein sequence it encodes; with less certainty (due to degeneracy in the code), a peptide sequence can be used to predict the approximate nucleic acid sequence that encodes it. Thus it has become feasible to go back and forth between the chemical languages of nucleic acids and proteins to obtain additional information about a polymer of one type or the other (Figure 6-37).

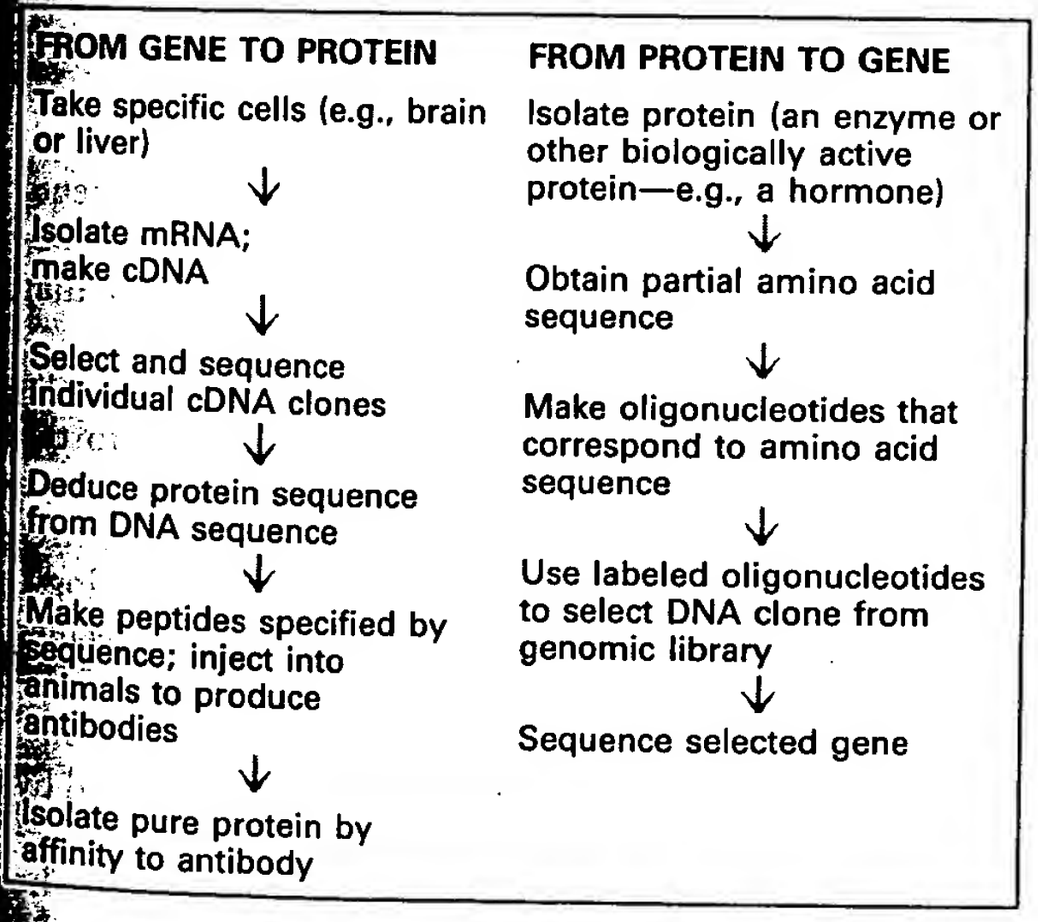
For example, if an mRNA region for a protein that is not yet isolated is cloned and sequenced, a synthetic peptide that is part of the protein can be prepared and used to provoke an antibody that will react with a protein containing that peptide. With such an antibody, the previously unisolated protein corresponding to the already isolated RNA can be identified in cells and purified. A reciprocal selection is also possible: if a protein has been purified and a short region of peptide sequence is avail-

◀ **Figure 6-36** In vitro mutagenesis: constructing DNA deletions and point mutations through the use of recombinant DNA techniques. (a) Deletions are made in cloned DNA in a plasmid by removing entire sections of DNA between two restriction sites or by cutting at a single restriction site and using the exonuclease Bal 31, which removes nucleotides from both ends of a cut double-stranded DNA molecule. Deletions of various lengths are chosen from a collection of such truncated molecules. (b) The two strands of a cloned DNA are separated, and a chemically synthesized oligonucleotide primer (see Figure 6-39) that is mismatched at a desired site is hybridized to one of the DNA strands and then extended by a DNA polymerase. Each strand of the new double-stranded molecule is copied during replication to produce a mixed population of the original DNA and mutants, which are then separately cloned.

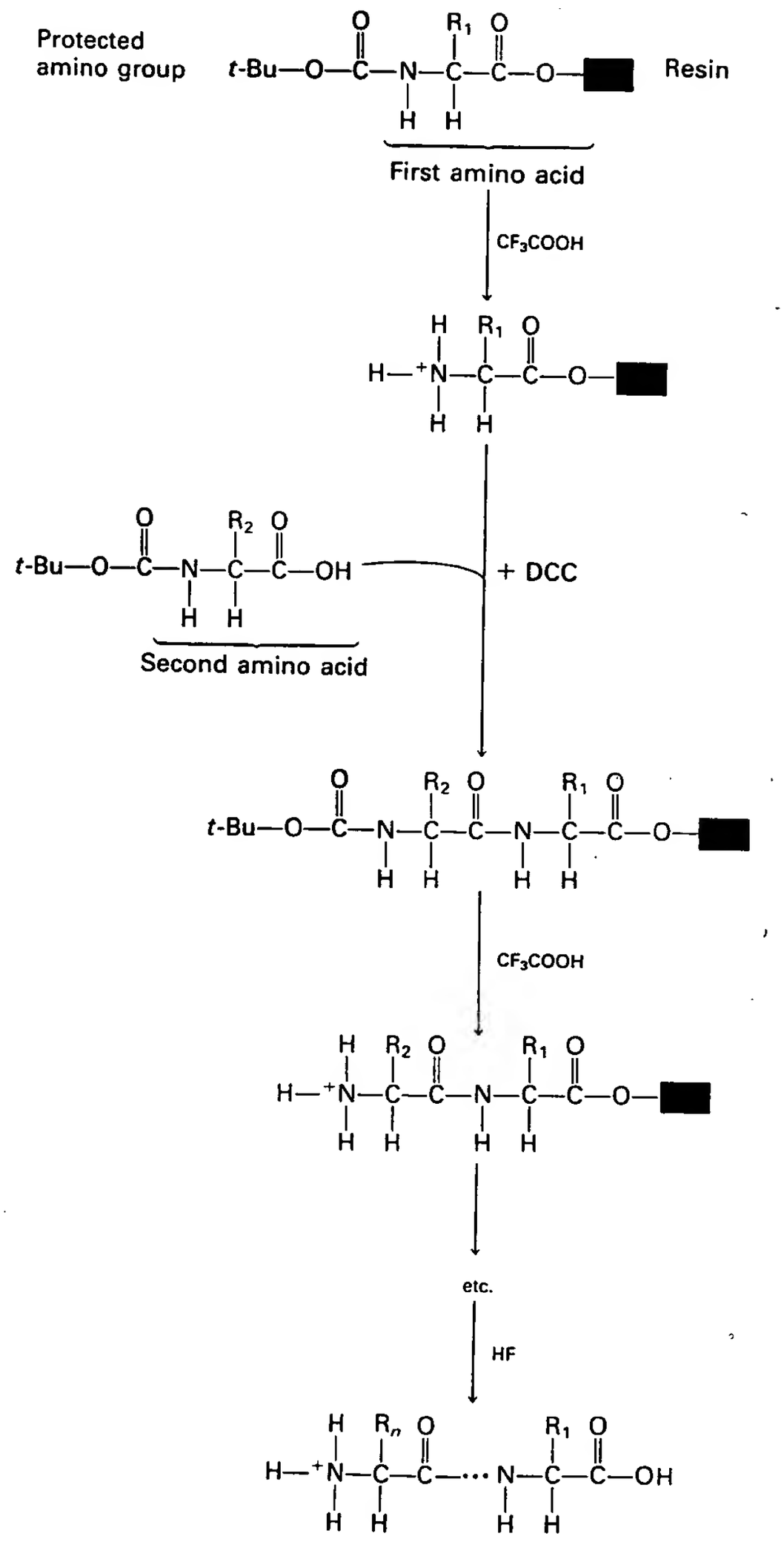
able, then oligonucleotides coding for that amino acid sequence can be synthesized and used to screen a genomic or cDNA library for the particular DNA sequence.

The degeneracy of the genetic code is an important consideration in choosing peptides from which to reconstruct partial mRNA sequences. For example, peptides containing arginine, leucine, or serine (six codons each) are to be avoided if possible. The best amino acids for making such probes are tryptophan and methionine (one codon apiece) and phenylalanine, tyrosine, histidine, aspartic acid, glutamic acid, asparagine, and glutamine (two codons each). The number of oligonucleotides that have to be synthesized to be certain of a perfect match with the native mRNA is multiplicative; for example, if a probe is to represent six amino acids with a total of 2, 3, 2, 1, 2, 2 codons, then 48 separate sequences are required.

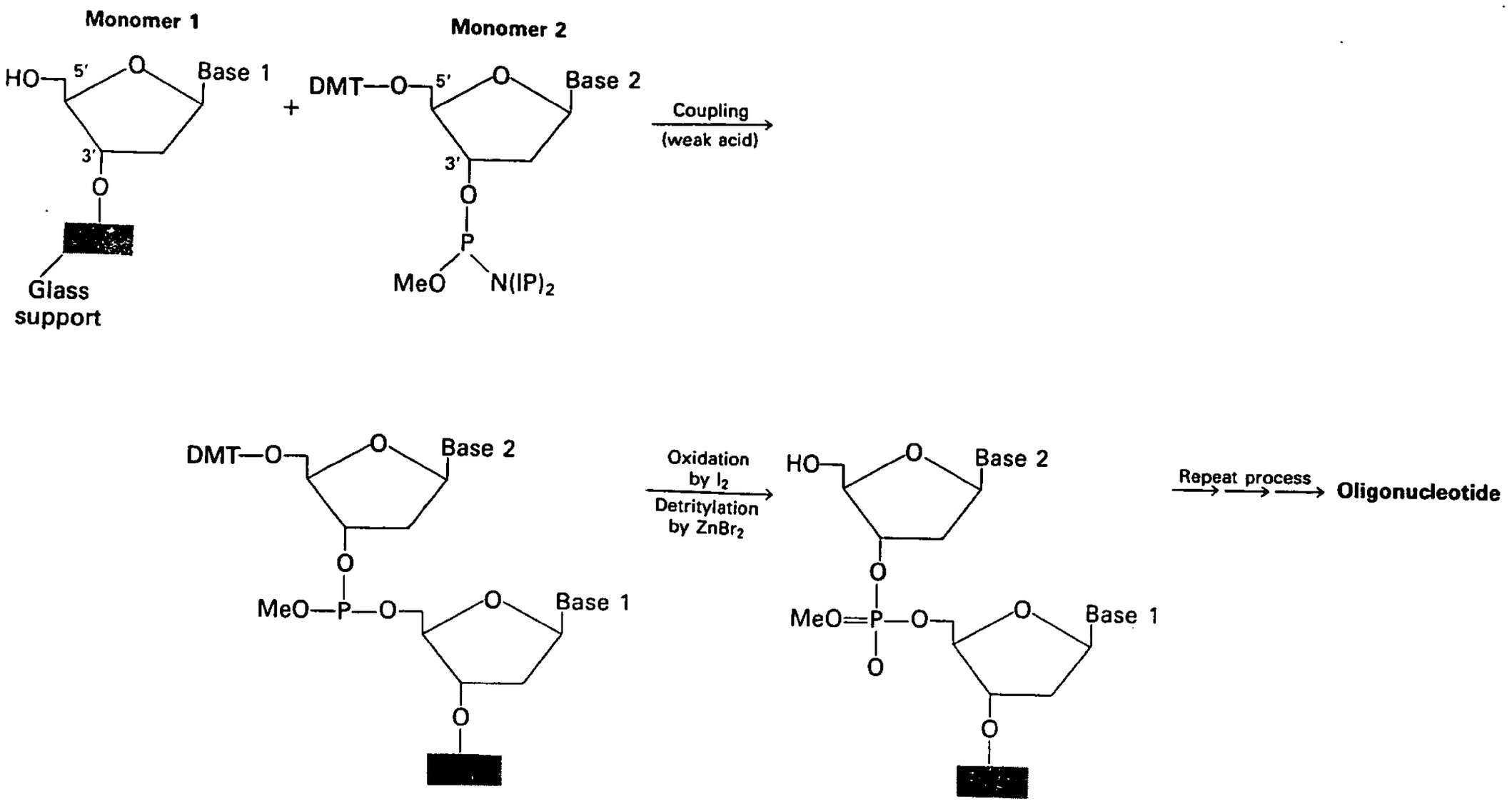
Techniques for the chemical synthesis of peptides (Figure 6-38) have been available for some time; techniques for DNA oligonucleotide synthesis (Figure 6-39) are also in wide use. The basic logic of these techniques is similar, although the chemistry is different. Note that during chemical synthesis, peptide chains grow from the carboxyl terminus to the amino terminus and DNA chains grow from the 3′ to the 5′ end. Both directions are opposite to the directions of biosynthetic reactions in cells or cell extracts.



▲ **Figure 6-37** It is now possible to identify an mRNA of interest (say, an mRNA present in only one part of the brain) and to use it to isolate the protein it encodes without knowing the function of that protein. On the other hand, it is possible to sequence part of a protein that has a specific function (say, an enzyme or a growth factor) and then to synthesize an oligonucleotide that can be used to identify and isolate the gene that encodes the complete protein.



▲ **Figure 6-38** Solid-phase peptide synthesis. The first amino acid of the desired peptide is attached at its carboxyl end by esterification to a resin. The amino group of the first amino acid in the peptide under construction is blocked by the attachment of a *tert*-butyloxycarbonyl group (yellow), which is removed by treatment with trifluoroacetic acid (CF_3COOH). The resulting free amino group forms a peptide bond with a second amino acid, which is presented with a reactive carboxyl group and a blocked amino group, together with the coupling agent dicyclohexylcarbodiimide (DCC). The process is repeated until the desired product is obtained; the peptide is then chemically cleaved with hydrofluoric acid (HF) from the resin. [See R. B. Merrifield, L. D. Vizioli, and H. G. Boman, 1982, *Biochemistry* 21:5020.]



▲ **Figure 6-39** Synthesis of oligonucleotides. The first nucleotide (monomer 1) is bound to a glass support by its 3′ hydroxyl; its 5′ hydroxyl remains available. The synthesis of the first internucleotide link is carried out by mixing monomer 1 with monomer 2, which contains a reactive 3′-diisopropyl phosphoramidite [$(IP)_2$] with attached methyl group (Me), a nucleotide derivative that has the blocking group 4′,4′-dimethoxytrityl (DMT) bound to its 5′ hydroxyl. In the presence of a weak acid, the two nucleotides couple to form a phosphodiester with phorphorus in a trivalent state. Oxidation by iodine (I_2) yields a phosphotriester in which the P is pentavalent; detritylation with zinc bromide ($ZnBr_2$) is carried out, and the process is repeated. The methyl groups on the phosphates are all removed at alkaline pH when synthesis is finished. [See S. L. Beaucage and M. H. Caruthers, 1981. *Tetrahedron Letters* 22:1859.]

Summary

An indispensable adjunct of modern molecular cell biology is the use of isotopes to label biologically important molecules. The isotopes may be radioactive (most commonly used are ^{3}H, ^{14}C, and ^{32}P) or density-labeled (for example, ^{15}N or ^{13}C). These tracers are widely used in cell-free biochemical experiments and in the observation of metabolic events within cells. Important considerations in the use of isotopes include the energy of the emitted particle during radioactive decay, the speed at which various labeled macromolecular precursors enter the cell, and the extent of exchange between compounds in the cell and the medium. For example, tritiated (^{3}H) compounds give the best autoradiographic images because the emitted β particle has a low energy and the image on the photographic emulsion is better defined.

Pulse-chase experiments using labeled amino acids or thymidine to study the synthesis of proteins or DNA can produce clear results because amino acids are exchanged between the cell and the medium within a minute or two; the thymidine enters a very small pool that is quickly consumed by cell growth. However, pulse-chase experiments with labeled RNA precursors are much less effective because ribonucleosides enter a large intracellular pool that is slowly consumed.

Techniques for separating purified molecules from cells have reached the level of a high art. In addition to the many varieties of chromatographic procedures, two basic methods—centrifugation and electrophoresis—are frequently applied to problems in molecular cell biology. Both techniques are most useful in separating molecules according to chain length. Separations of very large molecules that differ by less than 1 percent in size are routine. In addition, separation in two dimensions (by size and by charge) allows the total protein content of cells to be resolved into more than 5000 individual components. The use of electrophoresis to separate nucleic acids on the basis of size has become one of the most common laboratory procedures. In mixtures of chains of 500 nucleotides or less, chains of every length can be separated. These nucleic acid fragments can now be sequenced with such facility that DNA stretches thousands of nucleotides long are typically sequenced within days. Protein sequencing

of shorter peptides has been entirely automated, as has the chemical synthesis of oligonucleotides and peptides of 50 units or more in length.

Two aspects of nucleic acid biochemistry—molecular hybridization and nucleic acid enzymology—used in conjunction with microbial genetics have spawned an array of revolutionary techniques for identifying, cloning, and producing natural and mutant nucleic acid sequences. Molecular hybridization (both RNA-DNA and DNA-DNA), the fundamental method of testing the identity of a nucleic acid sample, underlies many of these applications. The detection of a single gene representing perhaps as little as one part in 10^6 of the total human genome is routinely carried out by a hybridization procedure known as the Southern blot. Especially sensitive are the Northern blot, which detects specific mRNA, and the Western blot, which employs antibodies to detect individual proteins.

Among the most important discoveries that allowed gene cloning was the recognition of the restriction endonucleases that cut DNA at characteristic restriction sites of 4–8 bp, thereby generating reproducible fragments from any genome. Enzymes that synthesize DNA and RNA are widely available in highly purified forms, as are enzymes that add to or remove nucleotides from the ends of nucleic acids and enzymes that join DNA segments. The clever use of these enzymes coupled with a deep understanding of microbial genetics that provides exquisitely designed selectable *vectors* to receive tailor-made pieces of DNA has made recombinant DNA experiments commonplace. Synthetic oligonucleotides allow planned deletion and mutation of genes by substitution of sequences in recombinant DNA. Today, any gene can be purified and the functional regions of its DNA sequences can be explored by reintroducing the DNA into cells and into whole organisms. As subsequent chapters will show these fantastic techniques have completely reshaped the way biology is carried out today.

References

Radioisotopes: The Indispensable Modern Means of Following Biological Activity

FREEMAN, L. M., and M. D. BLAUFOX. 1975. *Radioimmunoassay.* Grune & Stratton.

HENDEE, W. R. 1973. *Radioactive Isotopes in Biological Research.* Wiley.

Autoradiography

HUTEN, E., A. KUROUVA, and W. J. GEHRING. 1984. Spatial distribution of transcriptase from the segmentation gene *fushi tarazu* during *Drosophila* embryonic development. *Cell* 37:833–841.

PARDUE, M. L., and J. G. GALL. 1969. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Nat'l Acad. Sci. USA* 64:600–604.

ROGERS, A. W. 1979. *Techniques of Autoradiography,* 3d ed. Elsevier/North-Holland.

Labeling Cells and Pools

DINTZIS, H. 1961. Assembly of the peptide chains of hemoglobin. *Proc. Nat'l Acad. Sci. USA* 47:247–261.

EAGLE, H., and K. A. PIEZ. 1958. The free amino acid pool of cultured human cells. *J. Biol. Chem.* 231:533–545.

PUCKETT, L., and J. E. DARNELL JR. 1977. Essential factors in the kinetic analysis of RNA synthesis in HeLa cells. *J. Cell Phys.* 90:521–534.

Labeling Isolated Molecules

BOLTON, A. E., and W. M. HUNTER. 1973. The labeling of proteins to high specific radioactivities by conjugation to a ^{125}I-containing acylating agent. *Biochem. J.* 133:529–538.

RIGBY, P. W. J., M. DIECKMANN, C. RHODES, and P. BERG. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237–251.

Determining Sizes of Nucleic Acids and Proteins

CANTOR, C. R., and P. R. SCHIMMEL. 1980. *Biophysical Chemistry,* part II: *Techniques for the Study of Biological Structure and Function.* W. H. Freeman and Company.

DAWKINS, H. J. S., D. J. FERRIER, and T. L. SPENCER. 1987. Field inversion gel electrophoresis (FIGE) in vertical slabs as an improved method for large DNA separation. *Nuc. Acids Res.* 15:3634–3636.

SMITH, C. L., J. G. ECONOME, A. SCHUTT, S. KLCO, and C. R. CANTOR. 1987. A physical map of the *Escherichia coli* K12 genome. *Science* 236:1448–1453.

Centrifugation

DE DUVE, C. 1975. Exploring cells with a centrifuge. *Science* 189:186–194.

SCHEELER, P. 1980. *Centrifugation in Biology and Medical Science.* Wiley.

VINOGRAD, J. 1963. Sedimentation equilibrium in a buoyant density gradient. In *Methods in Enzymology,* vol. VI, S. P. Colowick and N. O. Kaplan, eds. Academic Press.

VINOGRAD, J., R. RASLOFF, and W. BAUER. 1967. A dye-buoyant density method for the detection and isolation of closed circular duplex DNA. *Proc. Nat'l Acad. Sci. USA* 57:1514–1521.

Electrophoresis

ANDREWS, A. T. 1986. *Electrophoresis, 2d ed.* Oxford University Press.

CANTOR, C. R., C. L. SMITH, and M. K. MATTHEW. 1988. Pulsed-field gel electrophoresis of very large DNA molecules. *Ann. Rev. Biophysics & Biophysical Chem.* 17:41–72.

CARLE, G. F., M. FRANK, and M. V. OLSON. 1986. Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. *Science* 232:65–70.

CELIS, J. E. and R. BRAVO. 1983. Two-dimensional gel electrophoresis of proteins. Academic Press.

CROTHERS, D. M. 1987. Gel electrophoresis of protein-DNA complexes. *Nature* 325:464–465.

MAIZEL, J. V. JR. 1971. Polyacrylamide gel electrophoresis of viral proteins. In *Methods in Virology*, vol. V, K. Maramorosch and H. Koprowski, eds. Academic Press.

O'FARRELL, P. H. 1975. High resolution, two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4007–4021.

RICKWOOD, D., and B. D. HAMES, eds. 1982. *Gel Electrophoresis of Nucleic Acids*. London: IRL Press Ltd.

Examining Sequences of Nucleic Acids and Proteins

Nucleic Acids: Hybridization

ALWINE, J. C., D. J. KEMP, and G. R. STARK. 1977. A method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Nat'l Acad. Sci. USA* 74:5350–5354.

BERK, A. J., and P. A. SHARP. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. *Cell* 12:721–732.

BRITTEN, R. J., and E. D. KOHNE. 1968. Repeated sequences in DNA. *Science* 161:529–540.

DAVIS, R. W., and N. DAVIDSON. 1968. Electron microscopic visualization of deletion mutations. *Proc. Nat'l Acad. Sci. USA* 60:243–250.

HALL, B. D., and S. SPIEGELMAN. 1961. Sequence complementarity of T2-DNA and T2-specific RNA. *Proc. Nat'l Acad. Sci. USA* 47:137–146.

HU, N., and J. MESSING. 1982. The making of single-stranded probes. *Gene* 17:271–277.

THOMAS, M., R. L. WHITE, and R. W. DAVIS. 1976. Hybridization of RNA to double-stranded DNA: formation of R loops. *Proc. Nat'l Acad. Sci. USA* 73:2294–2298.

THOMAS, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Nat'l Acad. Sci. USA* 77:5201–5205.

WETMUR, J. G., and N. DAVIDSON. 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.* 31:349–370.

Nucleic Acids: Restriction Enzymes

DAVIES, K. E. ed., 1988. *Genome Analysis: A Practical Approach*. Oxford: IRL Press.

NATHANS, D., and H. O. SMITH. 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Ann. Rev. Biochem.* 44:273–293.

NELSON, M., and M. MCCLELLAND. 1989. The effect of site-specific methylation on restriction–modification enzymes. *Nuc. Acids Res.* 17(Suppl.):r389–r415.

ROBERTS, R. J. 1989. Restriction enzymes and their isoschizomers. *Nuc. Acids Res.* 17(Suppl.):r347–r388.

SMITH, H. O. 1970. Nucleotide sequence specificity of restriction endonucleases. *Science* 205:455–462.

SMITH, H. O., and M. L. BERNSTIEL. 1976. A simple method for DNA restriction-site mapping. *Nuc. Acids Res.* 3:2387–2398.

SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.

Nucleic Acids: Sequence Analysis

BROWNLEE, G. 1972. Determination of sequences in RNA. In *Laboratory Techniques in Biochemistry and Molecular Biology*, vol. 3, T. S. Work and E. Work, eds. Elsevier/North-Holland.

CHURCH, G. M., and S. KIEFFER-HIGGINS. 1988. Multiplex DNA sequencing. *Science* 240:185–188.

DOOLITTLE, R. F. 1986. *Of Urfs and Orfs: A Primer on How to Analyze Derived Amino Acid Sequences*. University Science Books.

ERLICH, H. E., R. A. GIBBS, and H. H. KAZAZIAN JR., eds. 1989. *Polymerase Chain Reaction*. Cold Spring Harbor Laboratory.

MAXAM, A. M., and W. GILBERT. 1977. A new method for sequencing DNA. *Proc. Nat'l Acad. Sci. USA* 74:560–564.

MAXAM, A. M., and W. GILBERT. 1980. Sequencing end-labeled DNA with base-specific DNA. In *Methods in Enzymology*, vol. 65, L. Grossman and K. Moldave, eds. Academic Press.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUICHI, G. T. HORN, K. B. MULLIS, and H. A. ERLICH. 1987. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491.

SANGER, F. 1981. Determination of nucleotide sequences in DNA. *Science* 214:1205–1210.

Proteins: Sequence Analysis

CLEVELAND, D. W., S. G. FISCHER, M. W. KIRSCHNER, and U. K. LAEMMLI. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* 252:1102–1106.

HUNKAPILLER, M. W., and L. E. HOOD. 1983. Protein sequence analysis: automated microsequencing. *Science* 219:650–659.

INGRAM, V. M. 1956. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* 178:792–794.

SANGER, F. 1952. The arrangement of amino acids in proteins. *Adv. Protein Chem.* 7:1–67. A description of original sequencing, fingerprinting, and partial-chain-analysis methods.

WALSH, K. A., L. H. ERICSSON, D. C. PARMELEE, and K. TITANI. 1981. Advances in protein sequencing. *Annu. Rev. Biochem.* 50:261–284. A review of up-to-date methods, including automated sequence analysis.

Recombinant DNA, Gene Cloning, and Mutagenesis

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, R. E. MOORE, D. D. SEIDMAN, J. G. SMITH, J. A. STRUHL, and K. STRUHL. 1987. *Current Protocols in Molecular Biology*. Wiley.

BENTON, W. D., and R. W. DAVIS. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science* 196:180–183.

BERG, P. 1981. Dissections and reconstructions of genes and chromosomes. *Science* 213:296–303.

BOTSTEIN, D., and D. SHORTLE. 1985. Strategies and applications of in vitro mutagenesis. *Science* 229:1193–1201.

COHEN, S. N., A. C. Y. CHANG, H. W. BOYER, and R. B. HELLING. 1973. Construction of biologically functional bacterial

plasmids in vitro. *Proc. Nat'l Acad. Sci. USA* 70:3240–3244.

COOKE, H. 1987. Cloning in yeast: An appropriate scale for mammalian genomes. *Trends Genet.* 3:173–175.

GRUNSTEIN, M., and D. S. HOGNESS. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Nat'l Acad. Sci. USA* 72:3961–3965.

HOHN, B. 1979. In vitro packaging of λ and cosmid DNA. In *Methods of Enzymology: Recombinant DNA*, vol. 68, R. Wu, ed. Academic Press.

OKAYAMA, H., and P. BERG. 1982. High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.* 2:161–170.

OLD, R. W., and S. B. PRIMROSE. 1985. *Principles of Gene Manipulation: An Introduction to Genetic Engineering*. Blackwell Scientific Publications.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. *Molecular Cloning*. Cold Spring Harbor Laboratory.

SINGH, H., J. H. LEBOWITZ, A. S. BALDWIN JR., and P. A. SHARP. 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition-site DNA. *Cell* 52:415–423.

SMITH, M. 1985. In vitro mutagenesis. *Ann. Rev. Genet.* 19:423–462.

THOMAS, K. R., and M. R. CAPECCHI. 1987. Site-directed mutagenesis by gene targeting in mouse-embryo-derived stem cells. *Cell* 51:503–512.

Synthetic Peptide and Nucleotide Sequences

Synthesis of Polypeptides

KENT, S. B. H. 1988. Chemical synthesis of peptides and proteins. *Ann. Rev. Biochem.* 57:957–990.

MERRIFIELD, B. 1986. Solid phase synthesis. *Science* 232:341–347.

PELHAM, H. R. B., and R. J. JACKSON. 1976. An efficient mRNA dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* 67:247–256. A description of the most frequently used in vitro technique for protein synthesis.

Synthesis of Oligonucleotides

CARUTHERS, M. H. 1985. Gene synthesis machines: DNA chemistry and its uses. *Science* 230:281–285.

ITAKURA, K., and A. D. RIGGS. 1980. Chemical DNA synthesis and recombinant DNA studies. *Science* 209:1401–1405.

ITAKURA, K., J. J. ROSSI, and R. B. WALLACE. 1984. Synthesis and use of synthetic oligonucleotides. *Ann. Rev. Biochem.* 53:323–356.